

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

| | Result No. | Score | Query Match % | Length | DB | ID | Description |
|--|-----------------|-------|---------------|--------|--------|----|----------------------------|
| Scoring table: | IDENTITY_NUC | 1 | 911 | 100.0 | 911 | 6 | AX350366 Sequence |
| Gapop 1.0_0 , Gapext 1.0 | | 2 | 759 | 83.3 | 1248 | 6 | AX552095 Sequence |
| Scoring table: | US-10-069-434-4 | 3 | 745.8 | 81.9 | 1023 | 6 | AK458348 Sequence |
| Title: US-10-069-434-4 | | 4 | 745.8 | 81.9 | 1237 | 9 | BC052602 Homo sapi |
| Perfect score: 911 | | 5 | 745.8 | 81.9 | 3564 | 6 | AX833515 Sequence |
| Sequence: 1 cggaaatccggctcgatgttcc.....ccccaaatgcccaaggaaaaa 911 | | 6 | 745.8 | 81.9 | 3564 | 9 | AK095314 Homo sapi |
| | | 7 | 740.6 | 81.3 | 828 | 6 | AX817267 Sequence |
| | | 8 | 728.8 | 80.0 | 833 | 6 | AX817269 Sequence |
| | | 9 | 590 | 64.8 | 1052 | 10 | AF231123 Mus muscu |
| | | 10 | 590 | 64.8 | 1082 | 10 | BC064050 Mus muscu |
| | | 11 | 575.2 | 63.1 | 267093 | 2 | AC109405 Rattus no |
| | | 12 | 506.4 | 55.6 | 266620 | 2 | AC112556 Rattus no |
| | | 13 | 506.4 | 55.6 | 276378 | 2 | AC110854 Rattus no |
| | | 14 | 493.2 | 54.1 | 688 | 6 | BD2765148 Human |
| | | 15 | 481.4 | 52.8 | 2403 | 9 | AK093753 Homo sapi |
| | | 16 | 341 | 37.4 | 1231 | 4 | AJ417908 Monodelph |
| | | 17 | 321.2 | 36.4 | 914 | 9 | X05114 Human CDNA |
| | | 18 | 321.2 | 36.4 | 2785 | 11 | M33987 Human carbo |
| | | 19 | 321.2 | 36.4 | 2785 | 11 | G2854 human carbo |
| | | 20 | 329.6 | 36.2 | 1307 | 9 | EC07890 Homo sapi |
| | | 21 | 328 | 36.0 | 810 | 9 | IU1622 Gorilla gor |
| | | 22 | 326.4 | 35.8 | 810 | 9 | COPCAIRO Pan troglod |
| | | 23 | 323.2 | 35.7 | 1125 | 9 | IU5082 Macaca nem |
| | | 24 | 321.6 | 35.3 | 1440 | 4 | IU42178 Ovis aries |
| | | 25 | 311.8 | 34.2 | 1224 | 10 | M32452 Mouse carbo |
| | | 26 | 310.2 | 34.1 | 1195 | 10 | BC011223 Mus muscu |
| | | 27 | 304.2 | 33.4 | 786 | 5 | AY125007 Lepisoste |
| | | 28 | 297.8 | 32.7 | 1976 | 5 | AB055617 Tribolodo |
| | | 29 | 296.2 | 32.5 | 1445 | 5 | ZI4957 G.domesticu |
| | | 30 | 294.8 | 32.4 | 1486 | 5 | GGCAATR X04810 Chicken mRN |
| | | 31 | 294.2 | 32.3 | 783 | 5 | AY307082 Oncorhynch |
| | | 32 | 291.4 | 32.0 | 1199 | 5 | XI2639 Chicken mRN |
| | | 33 | 289.4 | 31.8 | 1459 | 6 | AX401999 Sequence |
| | | 34 | 289.4 | 31.8 | 1459 | 6 | X58294 R.norvegicus |
| | | 35 | 284.8 | 31.3 | 605 | 6 | J95751 Sequence 17 |
| | | 36 | 284.8 | 31.3 | 605 | 6 | AR182211 Sequence |
| | | 37 | 283.6 | 31.1 | 881 | 4 | M10412 Rabbit eryt |
| | | 38 | 283.4 | 31.1 | 1490 | 10 | K00811 Mouse carbo |
| | | 39 | 283.4 | 31.1 | 1542 | 10 | BC055291 Mys muscu |
| | | 40 | 281.8 | 30.9 | 1574 | 5 | DR051177 Danio rerio |
| | | 41 | 280 | 30.7 | 780 | 6 | AR111448 Sequence |
| | | 42 | 280 | 30.7 | 864 | 6 | AR077867 Sequence |
| | | 43 | 280 | 30.7 | 867 | 6 | AR159858 Sequence |
| | | 44 | 280 | 30.7 | 867 | 6 | AR216143 Sequence |
| | | 45 | 280 | 30.7 | 1363 | 9 | M36532 Human carbo |

ALIGNMENTS

| RESULT | 1 | LOCUS | AX350366 | DEFINITION | Sequence 4 from Patent WO0200840. | 911 bp | DNA | linear | PAT 06-FEB-2002 |
|-----------|---|----------|---|------------|---|-------------|-----|--------|-----------------|
| | | VERSION | AX350366 | ACCESSION | AX350366.1 | GI:18616025 | | | |
| | | KEYWORDS | | SOURCE | Homo sapiens (human) | | | | |
| | | ORGANISM | | | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Buteleostomi; Mammalia; Buthidae; Primates; Catarrhini; Hominidae; Homo | | | | |
| REFERENCE | 1 | AUTHORS | Thornton, M., Ramkumar, J., Tribouley, C.M., Yue, H., Nguyen, D.B., Yao, M.G., Patterson, C., Gandhi, A.R., Burford, N., Thangavelu, K., Baughn, M.R. | | | | | | |

Pred. No. is the number of results predicted by chance to have a

| | | |
|--|---|--|
| TITLE | Human lyases | |
| JOURNAL | Patent : WO 0200840-A 4 03-JAN-2002; | |
| FEATURES | Incyte Genomics, Inc. (US) | |
| source | Location/Qualifiers | |
| | 1. .911 | |
| | /organism="Homo sapiens" | |
| | /mol_type="unassigned DNA" | |
| | /db_xref="Taxon:9616" | |
| | /note="Incyte ID No: 63183333CB1" | |
| ORIGIN | | |
| Query Match | Score 911; DB 6; Length 911; | |
| Best Local Similarity | 100.0% ; Pred. No. 5.2e-234; Mismatches 0; Indels 0; Gaps 0; | |
| Matches 911; Conservative 0; Mismatches 0; Indels 0; Gaps 0; | | |
| QY | 1 CGGAATTGGCTTCGAGTTCCACCCGAGGACCATCGAGCTGGCTCACGGTACCGC 60 | |
| Db | 1 CGGAATTGGCTTCGAGTTCCACCCGAGGACCATCGAGCTGGCTCACGGTACCGC 60 | |
| QY | 61 GAGCCAACGGCCATTCACTGGAGGAATTTCCTATGGCTATGGCTATCGCAA 120 | |
| Db | 61 GAGCCAACGGCCATTCACTGGAGGAATTTCCTATGGCTATGGCTATCGCAA 120 | |
| QY | 121 TCTCCAATTGGATAAACTAAAGGTAAATGACTCTCCGACACTAGT 180 | |
| Db | 121 TCTCCAATTGGATAAACTAAAGGTAAATGACTCTCCGACACTAGT 180 | |
| QY | 181 ATCAAGTATGCCAACAGCTCACTAAATCTCAAGCGGCCATTCTTCATGTT 240 | |
| Db | 181 ATCAAGTATGCCAACAGCTCACTAAATCTCAAGCGGCCATTCTTCATGTT 240 | |
| QY | 241 GACTTGATGACACAGAACAAATCACTGTTGCTCTACTGGAAACTAC 300 | |
| Db | 241 GACTTGATGACACAGAACAAATCACTGTTGCTCTACTGGAAACTAC 300 | |
| QY | 301 AGGTAGGGGGTTTACCTTAACGGGTCGCGTGTGCTCTACTGGAACTAC 360 | |
| Db | 301 AGGTAGGGGGTTTACCTTAACGGGTCGCGTGTGCTCTACTGGAACTAC 360 | |
| QY | 361 GTAGATGGAGCTGAGCTATGGCTCAGCTGTTTCACTGGAAATTCTGACAATAAC 420 | |
| Db | 361 GTAGATGGAGCTGAGCTATGGCTCAGCTGTTTCACTGGAAATTCTGACAATAAC 420 | |
| QY | 421 CCCAGCTTGTGAGGAGCTCATGACCCAGATGAGCTGGTCTGGGGTTTTA 480 | |
| Db | 421 CCCAGCTTGTGAGGAGCTCATGACCCAGATGAGCTGGTCTGGGGTTTTA 480 | |
| QY | 481 CAGATTGGTAACCAAACTGATGAACTGCAACAAAGTACTGACATTGGTCCATTAA 540 | |
| Db | 481 CAGATTGGTAACCAAACTGATGAACTGCAACAAAGTACTGACATTGGTCCATTAA 540 | |
| QY | 541 GAAAAGGGTAACCAAAACTGATGAACTGCAACAAAGTACTGACATTGGTCCATTCC 600 | |
| Db | 541 GAAAAGGGTAACCAAAACTGATGAACTGCAACAAAGTACTGACATTGGTCCATTCC 600 | |
| QY | 601 TGGGACTACTGGACATATCTGGTCTCTACAGTCCACCTCTTGAAGTGTCA 660 | |
| Db | 601 TGGGACTACTGGACATATCTGGTCTCTACAGTCCACCTCTTGAAGTGTCA 660 | |
| QY | 661 TGGATTGTTAACGAACTTAAACATGGCTCTACAGTCCACCTCTTGAAGTGTCA 720 | |
| Db | 661 TGGATTGTTAACGAACTTAAACATGGCTCTACAGTCCACCTCTTGAAGTGTCA 720 | |
| QY | 721 CTCCCTGTCACAGUGGAGGGTGAAGCGACCTTCTGAGTATGAGTCACCTCTGTC 780 | |
| Db | 721 CTCCCTGTCACAGUGGAGGGTGAAGCGACCTTCTGAGTATGAGTCACCTCTGTC 780 | |
| QY | 781 CCCAGCTGAGGGAGGTGATGACATCTGGCTTAATGCAACCTCCAACTCTGGACTCA 840 | |
| Db | 781 CCCAGCTGAGGGAGGTGATGACATCTGGCTTAATGCAACCTCCAACTCTGGACTCA 840 | |
| QY | 841 AGTGAACCTCCACCTCAGCTGAGGACTGCTCTGACCTGATGAACTGCTGATG 900 | |
| Db | 841 AGTGAACCTCCACCTCAGCTGAGGACTGCTCTGACCTGATGAACTGCTGATG 900 | |
| QY | 901 CAGGGGGAAAA 911 | |
| Db | 901 CAGGGGGAAAA 911 | |
| RESULT 2 | | |
| LOCUS | AX552095 | |
| DEFINITION | Sequence 4 from Patent WO0162927. | |
| VERSION | AX552095 | |
| KEYWORDS | GI:25896333 | |
| SOURCE | Homo sapiens (human) | |
| ORGANISM | Homo sapiens | |
| Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo. | | |
| REFERENCE | 1 Barville, S.C., Greenawalt,L.B., Lincoln,S.E., Stockdreher,T.K., Amshay, S., Chang,S.C., Chen,W., D'Sa,S.A., Dat,T.C., Liu,T.F., Rosen, B.H., Russo,F.D., Spiro,P.A., Bradley,D.L., Chen,E., Cohen,H.J., Daffo,A., Daniels,S.E., Dufour,G.E., Flores,V., Fong,W.T., Hodgson,D.M., Jackson,S.J., Jones,J.Y., Panzer,S., Roseberry,A.M., Shah,P., Wright,R.J., Yap,P.E., Dahl,C.R. and Hillman,J.L. | |
| AUTHORS | | |
| TITLE | Polypeptides and corresponding polynucleotides for diagnostics and therapeutics | |
| JOURNAL | Patent: WO 0162927-A 4 30-AUG-2001; | |
| FEATURES | Incyte Genomics, Inc. (US) | |
| source | 1. .1248 | |
| | /organism="Homo sapiens" | |
| | /mol_type="unassigned DNA" | |
| | /db_xref="taxon:9606" | |
| | /note="Incyte ID No: LI: 090574.1:20000FEB01" | |
| ORIGIN | | |
| Query Match | 83.3%; Score 759; DB 6; Length 1248; | |
| Best Local Similarity | 100.0% ; Pred. No. 4e-193; Matches 759; Conservative 0; Mismatches 0; Indels 0; Gaps 0; | |
| Matches | | |
| QY | 1 CGGAATTGGCTCGAGTTCCACCCGGAGGACCATGGTCTGGGGTCACTGGCTACCGC 60 | |
| Db | 1 CGGAATTGGCTCGAGTTCCACCCGGAGGACCATGGTCTGGGGTCACTGGCTACCGC 60 | |
| QY | 15 CGGAATTGGCTCGAGTTCCACCCGGAGGACCATGGTCTGGGGTCACTGGCTACCGC 74 | |
| Db | 15 CGGAATTGGCTCGAGTTCCACCCGGAGGACCATGGTCTGGGGTCACTGGCTACCGC 74 | |
| QY | 61 GAGCACACGGCTCTATTCAGTGAAGGAAATTCTGAGTATGGTATGTTGATGACGCCAA 120 | |
| Db | 61 GAGCACACGGCTCTATTCAGTGAAGGAAATTCTGAGTATGGTATGTTGATGACGCCAA 120 | |
| QY | 75 GAGCACACGGCTCTATTCAGTGAAGGAAATTCTGAGTATGGTATGACGCCAA 134 | |
| Db | 75 GAGCACACGGCTCTATTCAGTGAAGGAAATTCTGAGTATGGTATGACGCCAA 134 | |
| QY | 121 TCTCCATTGATGAGATAAACTGAGTAAATCATGAGAACTTCTCCGACCTACTGT 180 | |
| Db | 121 TCTCCATTGATGAGATAAACTGAGTAAATCATGAGAACTTCTCCGACCTACTGT 180 | |
| QY | 135 TCTCCATTGATGAGATAAACTGAGTAAATCATGAGAACTTCTCCGACCTACTGT 194 | |
| Db | 135 TCTCCATTGATGAGATAAACTGAGTAAATCATGAGAACTTCTCCGACCTACTGT 194 | |
| QY | 181 ATCAAGATGACCCAAAGCTCAGTAAATCATGAGAACTTCTCCGACCTACTGT 240 | |
| Db | 181 ATCAAGATGACCCAAAGCTCAGTAAATCATGAGAACTTCTCCGACCTACTGT 240 | |
| QY | 195 ATCAAGATGACCCAAAGCTCAGTAAATCATGAGAACTTCTCCGACCTACTGT 254 | |
| Db | 195 ATCAAGATGACCCAAAGCTCAGTAAATCATGAGAACTTCTCCGACCTACTGT 254 | |
| QY | 241 GACTTGTGACACAGAGAACAACTGAGTAAATCATGAGAACTTCTCCGACCTACTGT 300 | |
| Db | 241 GACTTGTGACACAGAGAACAACTGAGTAAATCATGAGAACTTCTCCGACCTACTGT 300 | |
| QY | 255 GACTTGTGACACAGAACAACTGAGTAAATCATGAGAACTTCTCCGACCTACTGT 314 | |
| Db | 255 GACTTGTGACACAGAACAACTGAGTAAATCATGAGAACTTCTCCGACCTACTGT 314 | |
| QY | 301 AGGTACGGCAGGGTCACTGGCTCTGGTCTGGGGTCTGAGTACGGCTCGGACACATA 360 | |
| Db | 301 AGGTACGGCAGGGTCACTGGCTCTGGTCTGGGGTCTGAGTACGGCTCGGACACATA 360 | |
| QY | 315 AGTTACGGCAGGGTCACTGGCTCTGGTCTGGGGTCTGAGTACGGCTCGGACACATA 374 | |
| Db | 315 AGTTACGGCAGGGTCACTGGCTCTGGTCTGGGGTCTGAGTACGGCTCGGACACATA 374 | |
| QY | 361 GTAGATGGAGTGGCTCTGGCTGAGGCTCCATGTTGCTGAGTACGGCTCTGGGGTCTGAGTAC 420 | |
| Db | 361 GTAGATGGAGTGGCTCTGGCTGAGGCTCCATGTTGCTGAGTACGGCTCTGGGGTCTGAGTAC 420 | |
| QY | 375 GTAGATGGAGTGGCTCCATGTTGCTGAGGCTCCATGTTGCTGAGTACGGCTCTGGGGTCTGAGTAC 434 | |
| Db | 375 GTAGATGGAGTGGCTCCATGTTGCTGAGGCTCCATGTTGCTGAGTACGGCTCTGGGGTCTGAGTAC 434 | |
| QY | 421 CCCAGCTTGTGACACAGAGAACAACTGAGTAAATCATGAGAACTTCTCCGACCTACTGT 480 | |
| Db | 421 CCCAGCTTGTGACACAGAGAACAACTGAGTAAATCATGAGAACTTCTCCGACCTACTGT 480 | |
| QY | 435 CCCAGCTTGTGACACAGAGAACAACTGAGTAAATCATGAGAACTTCTCCGACCTACTGT 494 | |
| Db | 435 CCCAGCTTGTGACACAGAGAACAACTGAGTAAATCATGAGAACTTCTCCGACCTACTGT 494 | |

| | | | | | | | |
|-----------------------|--|---|----------------|--------------|---------|---|---|
| Qy | 481 | CAGATGGTGAACCTAATTCCAACCTGCAAAAGATTACTGACACTTGGATTCCATTAA | 540 | Db | 317 | ACCTTACTCTGGGGTCGGCTATGACCACGGCTCCAGCACATAGTAGATGGATGGCT | 376 |
| Db | 495 | CAGATGGGACCTAATTCCAACCTGCAAAAGATTACTGACACTTGGATTCCATTAA | 554 | Db | 301 | ACCTTACTCTGGGGTCGGCTATGACCACGGCTCCAGCACATAGTAGATGGATGGCT | 360 |
| Qy | 541 | GAAAAGGTAAACAAACTGATTCAAAATTGACCTATGCTGCTCCACATCC | 600 | Qy | 377 | ATGCTGGAGGTCTCATGTTGTTCACTGGAAATTCAAGAACATACCCAGGTGAGG | 436 |
| Db | 555 | GAAAAGGTAAACAAACTGATTCAAAATTGACCTATGCTGCTCCACATCC | 614 | Db | 361 | ATGCTGAGGTCTCATGTTGTTCACTGGAAATTCAAGAACATACCCAGGTGAGG | 420 |
| Qy | 601 | TGGGACTACTGGACATATCTGGTTCTCATGTCACCTTGAGAGTGTACA | 660 | Qy | 437 | CAGCTCATGAAACGATGGACTGGCTCTGGGATGTTTACAGATTGGTGAACCTA | 496 |
| Db | 615 | TGGGACTACTGGACATATCTGGTTCTCATGTCACCTTGAGAGTGTACA | 674 | Db | 421 | CAGCTCATGAAACGATGGACTGGCTCTGGGATGTTTACAGATTGGTGAACCTA | 480 |
| Qy | 661 | TGGATGTTAAAGAACCTATAAACATCAGCTCTAACAGCTGCCAAATTGCCAGT | 720 | Qy | 497 | ATTCCCAACTGAAAGATTACTGACACTTGGATTCATTAAGAAAGGTAAACAA | 556 |
| Db | 675 | TGGATGTTAAAGAACCTATAAACATCAGCTCTAACAGCTGCCAAATTGCCAGT | 734 | Db | 481 | ATTCCCAACTGCAAAGATTACTGACACTTGGATTCATTAAGAAAGGTAAACAA | 540 |
| Qy | 721 | CTCCCTGCAAGCGAGGTTGAGCAGCAGCTTCTG | 759 | Qy | 557 | CTCGATTACACAAATTGGACCTATTGACCTTCTGCTTCACCATCTGGACTACTGGACAT | 616 |
| Db | 735 | CTCCCTGCAAGCGAGGTTGAGCAGCAGCTTCTG | 773 | Db | 541 | CTCGATTACACAAATTGGACCTATTGACCTTCTGCTTCACCATCTGGACTACTGGACAT | 600 |
| RESULT 3 | AX449348 | AX458348 | 1023 bp | DNA | linear | PAT 08-JUL-2002 | Qy |
| LOCUS | Sequence 19 from Patent WO2463385. | | | | | | 677 AACCTTAAACATCAGTCTCAAGCTGGCAAATTGCGAGTCTCTGTGCAAGCGG |
| DEFINITION | | AX458348 | | | | | 736 |
| VERSION | AX458348.1 | GI:21725020 | | | | | 661 AACCTTAAACATCAGTCTCAAGCTGGCAAATTGCGAGTCTCTGTGCAAGCGG |
| KEYWORDS | | | | | | | 720 |
| SOURCE | Homo sapiens (human) | | | | | | 737 ACGGTGAGGACCGAGCTTCTGATAGAGTCCTACTGTGCAAGCCAGGTGGAGGCA |
| ORGANISM | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; | | | | | | 796 |
| | Mammalia; Eutheria; Primates; | | | | | | Db 721 ACGGTGAGGACCGAGCTTCTGATAGAGTCCTACTGTGCAAGCCAGGTGGAGGCA |
| | 1 Tang, Y. T., Griffin, J.A., Yue, H., Lee, E. A., Baughn M. R., | | | | | | 780 |
| REFERENCE | Gaudhi, B.M., Wallia, N.K., Lee, S., Ramkumar, J., Warren, B.A., | | | | | | |
| AUTHORS | Gandhi, A.R., Lu, D.A., Lu, Y., Yao, M.G., Ding, L., Troubouley, C.M., | | | | | | |
| | Sacjanawala, M.M., Arvizu, C. and Hillman, J.L. | | | | | | |
| | Enzymes | | | | | | |
| | Patent: WO 02463385-A 19 13-JUN-2002; | | | | | | |
| JOURNAL | Incyte Genomics, Inc. (US) | | | | | | |
| FEATURES | Location Qualifiers | | | | | | |
| | 1..1023 | | | | | | |
| | /organism="Homo sapiens" | | | | | | |
| | /mol_type="unassigned DNA" | | | | | | |
| | /db_xref="taxon:9606" | | | | | | |
| | /note="Incyte ID No: 8159895CB1" | | | | | | |
| ORIGIN | | | | | | | |
| Query Match | 81.9% | Score 745.8; | DB 6; | Length 1023; | | | |
| Best Local Similarity | 97.2%; | Prev. No. 1.4e-189; | | | | | |
| Matches 759; | Conservative | 0; | Mismatches 22; | Indels 0; | Gaps 0; | | |
| Qy | 17 | TTCCACCCGGGGACCTTGCGAGGCATGGTGGGATACCGGAGCAACGGCTTA | 76 | Db | 1 | TTCCACCCGGGGACCTTGCGAGGCATGGTGGGATACCGGAGCAACGGCTTA | 60 |
| | 17 | TTCCACCCGGGGACCTTGCGAGGCATGGTGGGATACCGGAGCAACGGCTTA | 76 | Db | 77 | TTCACTGGAGGAATTTCCTATTGCTGATGGTATCAGAACATTGAGATA | 136 |
| | 17 | TTCCACCCGGGGACCTTGCGAGGCATGGTGGGATACCGGAGCAACGGCTTA | 76 | Db | 61 | TTCACTGGAGGAATTTCCTATTGCTGATGGTATCAGAACATTGAGATA | 120 |
| | 137 | AAACCAAGAAGTGAATATGACTCTCCCTCCGACCACTTAGTCAAGTATGCCAA | 196 | Qy | 197 | GTCAGCTAAATCATGCAAGGAAATTTCCTATTGCTGATGGTATCAGAACAG | 256 |
| | 137 | AAACCAAGAAGTGAATATGACTCTCCCTCCGACCACTTAGTCAAGTATGCCAA | 196 | Qy | 181 | GCTCAGTAAATCATGCAAGGAAATTTCCTATTGCTGATGGTATCAGAACAG | 240 |
| | 121 | AAACCAAGAAGTGAATATGACTCTCCCTCCGACCACTTAGTCAAGTATGCCAA | 180 | Db | 257 | AGAACAAATCAGTCTGGTGTGTCAGCTGAAAGCTACAGTCTGGCAGGTT | 316 |
| | 121 | AAACCAAGAAGTGAATATGACTCTCCCTCCGACCACTTAGTCAAGTATGCCAA | 180 | Db | 241 | AGAACAAATCAGTCTGGTGTGTCAGCTGAAAGCTACAGTCTGGCAGGTT | 300 |
| TITLE | | | | | | | |
| JOURNAL | Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002) | | | | | | |
| MEDIUM | 22387922 | | | | | | |
| PRIMED | 12447922 | | | | | | |

| | | | |
|--|--|---|--|
| REFERENCE | 2 (bases 1 to 1237) | Qy | 77 TTCACTCGAAGGAATTTCCTATTTGCTGATGGATCAGCAAATTCACAAATTGAGATA 136 |
| AUTHORS | Strausberg,R. | Db | 346 TTCACTCGAAGGAATTTCCTATTTGCTGATGGATCAGCAAATTGAGATA 405 |
| TITLE | Direct Submission | Qy | 137 AAACCAAAAGTGAATATGACTCTCCGACCACTTAGTATAAGTGAACCA 196 |
| JOURNAL | Submitted (16-MAY-2003) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA | Db | 406 AAACCAAAAGTGAATATGACTCTCCGACCACTTAGTATAAGTGAACCA 465 |
| REMARK | NIH-MGC Project URL: http://mgc.nci.nih.gov | Qy | 197 GCTAGCTTAATCATCGAACAGCGCCATTCTCTCAATGTTGACTTTGATGACAG 256 |
| COMMENT | Contact: MGCB help desk Email: cgabps-r@mail.nih.gov | Db | 466 GCTAGCTTAATCATCGAACAGCGCCATTCTCTCAATGTTGACTTTGATGACAG 525 |
| Tissue Procurement: | ATCC | Qy | 257 AGAACAAATCAGTTCTCGGTCTCACTGAAAGTACAGCTTACGGAGGTT 316 |
| CDNA Library Preparation: | Rubin Laboratory | Db | 526 AGAACAAATCAGTTCTCGGTCTCACTGAAAGTACAGCTTACGGAGGTT 585 |
| CDNA Library Arrayed by: | The I.M.A.G.E. Consortium (LILNL) | Qy | 586 ACCTTCACTGGGGTCCGCTGATGACCAACAGCGCCATTCTCTCAATGTTGAGCT 645 |
| DNA Sequencing by: | National Institutes of Health Intramural Sequencing Center (NISC), Gaithersburg, Maryland; | Qy | 317 ACCTTCACTGGGGTCCGCTGATGACCAACAGCGCCATTCTCTCAATGTTGAGCT 376 |
| Web site: | http://www.nisc.nih.gov/ | Db | 377 ATGCTGAGAGTCATGTTGACTGAAATTCAGCAAATACCCAGCTTGTGAGG 436 |
| Akhter,N., Ayele,K., Beckstrom-Sternberg,S.M., Benjamin,B., Blakesley,R.W., Bouffard,G.G., Breen,K., Brinkley,C., Brooks,S., Deitchman,N.L., Grainger,X., Guan,X., Gupta,J., Hadiighi,P., Hansen,N., Ho,S.-L., Karins,E., Kwong,P., Laric,P., Legaspi,R., Maduro,O.L., Masiello,C., Mastriani,B., McCloskey,J.C., McDowell,J.J., Pearson,R., Scantripoli,S., Thomas,P.J., Touchman,J.W., Tsugeon,C., Vogt,J.L., Walker,M.A., Wetherby,K.D., Wiggins,L., Young,A., Zhang,L.-H. and Green,E.D. | Qy | 646 ATGCTGAGAGTCATGTTGACTGAAATACCCAGCTTGTGAGG 705 | |
| Clones distribution: | MG C clone distribution information can be found through the I.M.A.G.E. Consortium/LILNL at: http://image.lilnl.gov | Qy | 437 CAGCTCATGAAACGAGTGGACTGGCTCTGGAGTGTTTACAGATTTGAACTTA 496 |
| Series: | IATL Plate: 7 Row: F Column: 12 | Db | 706 CAGCTCATGAAACGAGTGGACTGGCTCTGGAGTGTTTACAGATTTGAACTTA 765 |
| This clone was selected for full length sequencing because it passed the following selection criteria: Hexamer frequency ORF analysis, GenomeScan gene prediction, Similarity but not identity to protein. | Location/Qualifiers | Qy | 497 ATTCCCAACTGCAAAAGTAACTGACACTTTGATTTCCATTAAGAAAAGGTTAACAA 556 |
| FEATURES | source | Db | 766 ATTCCCAACTGCAAAAGTAACTGACACTTTGATTTCCACATTTGAAAGAAAGGTTAACAA 825 |
| /organism="Homo sapiens" | | Qy | 557 CTGATTCAAAATTGACCTATGCTCTGCTCTGGACTACTGGACAT 616 |
| /mol_type="mRNA" | | Db | 826 CTCGATTCACAAATTGACCTATGCTCTGCTCTGGACTACTGGACAT 885 |
| /db_xref="BACAXON_9606" | | Qy | 617 ATCTGGTTCTCATGTTACAGTTCCACCTCTCTGAGGTGTCACATGGATTGTTAAAGC 676 |
| /clone="MGC159868 IMAGE:6083105" | | Db | 886 ATCTGGTTCTCATGTTACAGTTCCACCTCTCTGAGGTGTCACATGGATTGTTAAAGC 945 |
| /tissue_type="Pancreas, ductal carcinoma" | | Qy | 677 AACCTATAAACATCAGCTCTCACAGTGGCCAATTTCGCAGTCCTCTGTGCAAGCGGG 736 |
| /clone_id="NINH_MGC_1110" | | Db | 946 AACCTATAAACATCAGCTCTCACAGTGGCCAATTTCGCAGTCCTCTGTGCAAGCGGG 1005 |
| /lab_host="DH10B-R" | | Qy | 737 AGGTGAAGCAGGCGCTTTCTGTGATAGTGTCACTCTGTACACCGGCTGAGGCCA 796 |
| /note="Vector: POTB7" | | Db | 1006 AGGTGAAGCAGGCGCTTTCTGTGATAGTGTCACTCTGTACACCGGCTGAGGCCA 1065 |
| gene | | Qy | 797 G 797 |
| 1..1237 | | Db | 1066 G 1066 |
| /gene="MGC59868" | | RESULT 5 | |
| /note="synonym: FLJ37995" | | AX833515 | |
| /db_xref="LocusID:377677" | | LOCUS | 3564 bp |
| 303..1091 | | DEFINITION | Sequence 639 from Patent EP1347046. |
| /codon_start=1 | | ACCESSION | AX833515 |
| /product="similar to carbonic anhydrase 1.3" | | VERSION | AX833515.1 GI:3.9919650 |
| /protein_id="AAH52102.1" | | KEYWORDS | |
| /db_xref="GI:30851209" | | SOURCE | Homo sapiens (human) |
| /translition="MSRLSLWGRHEHNGPINKKEFPFLADGDDQSPIEIKTKEVKYDSSDRPLSLKDPSSAKILNSGHSFNDYVWPSDFDDTSVRSVGPGLRQLQEVHNSQADITDPLSIKERSKQTKRTNFEDLSLPPSWDWTYCSLTVPLLESVTWVLKQPINISSQQLAKFRSLRSTCTAGEAAFLVSNRPPPLGRKVRSFSH" | | ORGANISM | Homo sapiens |
| missc_feature | | ORIGIN | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. |
| Query Match 81.9%; Score 745.8; DB 9; Length 1237; Best Local Similarity 97.2%; Pred. No. 1.4e-189; Indels 0; Gaps 0; Matches 759; Conservative | | REFERENCE | 1 Isogai,T., Sugiyama,T., Otsuki,T., Wakamatsu,A., Sato,H., Ishii,S., Yamamoto,J.I., Isono,Y., Hio,Y., Otsuka,K., Nagai,K., Irie,R., Tamechika,I., Seki,N., Yoshihikawa,T., Otsuka,M., Nagahari,K. and Masuho,Y. |
| Db | | AUTHORS | Full-length cDNA sequences |
| 17 TTCCACCCGGAGGACCATGATGGCTAGCTGGGATAACGGTCTTA 76 | | TITLE | Patent: EP 1347046-A 639 24-SEP-2003; |
| 286 TTCCACCCGGAGGACCATGATGGCTAGCTGGGATAACGGTCTTA 345 | | JOURNAL | Research Association for Biotechnology (JP) |
| | | FEATURES | Location/Qualifiers |

| ORIGIN | source | Query Match Best Local Similarity Matches | Score 91.9% ; 97.2% ; 759 ; | Length 3564 ; 3564 ; Indels 0 ; | Gaps 0 ; | KEYWORDS Homo sapiens unassigned DNA | VERSION AK095314.1 | GI:21754548 oligo capping; fis (full insert sequence). |
|------------|--------|---|--|--|----------------------|--|-----------------------|---|
| | | /db_xref="taxon:9606" | | | | Homo sapiens (human) | SOURCE Organism | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo |
| QY | 17 | TTCACCCGAGGACATGGAGACTACTGGGATACCGCAGAACAGGTCCCA | 76 | Shinjima, T., Irie, R., Otsuka, T., Saito, H., Wakanatsu, A., Ishii, S., Supijima, F., Wakebe, H., Ono, T., Hischigaki, H., Watanabe, T., Ozaki, K., Yamamoto, J., Isogo, Y., Kawai-Hio, Y., Saito, K., Nishikawa, T., Kumura, K., Yamashita, H., Matsushita, K., Nakamura, Y., Sekine, M., Kiuchi, H., Kanda, K., Wagatsuma, M., Murakawa, K., Kanehori, K., Takahashi-Fujii, A., Oshima, A., Sugiyama, A., Kawakami, B., Nagano, S., Nagahara, K., Masuno, Y., Nagai, K., and Isogai, T. | REFERENCE AUTHORS | Tanigami, A., Fujiwara, T., Shibahara, T., Goto, Y., Hirao, M., Hirao, M., Hischigaki, H., Ono, T., Hischigaki, H., Watanabe, T., Ozaki, K., Supijima, T., Irie, R., Otsuka, T., Saito, H., Wakanatsu, A., Ishii, S., Yamamoto, J., Isogo, Y., Kawai-Hio, Y., Saito, K., Nishikawa, T., Kumura, K., Yamashita, H., Matsushita, K., Nakamura, Y., Sekine, M., Kiuchi, H., Kanda, K., Wagatsuma, M., Murakawa, K., Kanehori, K., Takahashi-Fujii, A., Oshima, A., Sugiyama, A., Kawakami, B., Suzuki, Y., NEDO human cDNA sequencing project | | |
| Db | 35 | TTCCACCCGAGGACCATGGGGATACCGCAGAACAGGTCCCA | 94 | | | | | Unpublished 2 (bases 1 to 3564) Isogai, T. and Yamamoto, J. |
| QY | 77 | TTACTGGAAAGAATTTCCTTATGCTGATGGTATCAGCAATTCATTGAGATA | 136 | | | | | Direct Submission Submitted (04-JUL-2002) Takao Isogai, FLJ Project (HRI Team); 2-6-7 |
| Db | 95 | TTACTGGAAAGAATTTCCTTATGCTGATGGTATCAGCAATTCATTGAGATA | 154 | | | | | Kazusa-Kamatari, Kisarazu, Chiba 292-0812, Japan (E-mail: genomics@tri.co.jp, Tel: 81-438-52-3975, Fax: 81-438-52-3986) |
| QY | 137 | AAACCAAAGAAGTGAATAATGACTCTTCCCGACCACTTAGTCAAGATGACCAA | 196 | | | | | NEDO human cDNA sequencing Project supported by Ministry of Economy, Trade and Industry of Japan; cDNA full insert sequencing: |
| Db | 155 | AAACCAAAGAAGTGAATAATGACTCTTCCCGACCACTTAGTCAAGATGACCAA | 214 | | | | | Research Association for Biotechnology (RAB); cDNA library construction; Helix Research Center etc.); 5' - & 3'-end one pass sequencing: RAB, HRI, and Biotechnology Center, National Institute of Technology and Evaluation; clone selection for full insert sequencing: HRI and RAB. |
| QY | 197 | GCTCAGCTAAATCATCAGGAAACAGGGCATTCCCTCAATGTTGACTTTSATGACACAG | 256 | | | | | Location/Qualifiers 1. .3564 /organism="Homo sapiens" /mol_type="mRNA" /db_xref="taxon:9606" /clone="CTONG2011855" /tissue_type="tongue, tumor tissue" /clone_id="CTONG21" /note="cloning vector: pME18SFL3" /note="unnamed protein product" /codon_start=1 /protein_id="PAC04528-1" /db_xref="GI:21754549" /translation:"MSRLWGTYRHNGP IHWKEFPIANGDQQSPLIEITKVEKYDSSLRPSIKYDPESSAKLISNGHSFNDDETDNTSVLRGGSPFTSYRLQFHGNPSADDHGSHEHIVCVSYAAELHVYHNSDKYKPFVVAHEPDGLAVLGFLQGEPSNLDLSSLLPPSDWYTFPSLTVPLLESVTWVLKOPINISSQQLAKRSLLCTAEGAFAAFLVSNHRPPQPLGRKTRASPH" |
| Db | 215 | GCTCAGCTAAATCATCAGGAAACAGGGCATTCCCTCAATGTTGACTTTSATGACACAG | 274 | | | | | |
| QY | 257 | AGACAATATGTTCTCGGTGTGGCTCTAACTGAAAGTACAGTAACTGGAGCTTC | 316 | | | | | |
| Db | 275 | AGACAATATCAGTTCTCGGTGTGGCTCTAACTGAAAGTACAGTAACTGGAGCTTC | 334 | | | | | |
| QY | 317 | ACCTTCACTGGGGTGGGTATGACCAACGGTCCAGAACATAGTAGATGGAGTAGCT | 376 | | | | | |
| Db | 335 | ACCTTCACTGGGGTGGGTGGGTATGACCAACGGTCCAGAACATAGTAGATGGAGTAGCT | 394 | | | | | |
| QY | 377 | ATGCTGCAAGACTCCATGTTCTCACTGAAATTTCAGCAAAATACCCAGCTTGTGAGG | 436 | | | | | |
| Db | 395 | ATGCTGCAAGACTCCATGTTCTCACTGAAATTTCAGCAAAATACCCAGCTTGTGAGG | 454 | | | | | |
| QY | 437 | CAGCTCATGAAACAGATGGACTGGCTCTGGCTCTGGAGTTACAGATGGTGCCTCA | 496 | | | | | |
| Db | 455 | CAGCTCATGAAACAGATGGACTGGCTCTGGCTCTGGAGTTACAGATGGTGCCTCA | 514 | | | | | |
| QY | 497 | ATCCCCAACTGAAAGATTAATGACACTGGATCCATTAAGAAAGGTTAACAA | 556 | | | | | |
| Db | 515 | ATCCCCAACTGAAAGATTAATGACACTGGATCCATTAAGAAAGGTTAACAA | 574 | | | | | |
| QY | 557 | CTCGATTCACAATTTGACCTATTGTCTCTCTTCCACACATCCTGGAACTTGAGCAT | 616 | | | | | |
| Db | 575 | CTCGATTCACAATTTGACCTATTGTCTCTTGAGCAT | 634 | | | | | |
| QY | 617 | AATCTGGTTCTCTTACAGTTCACTTCAACCTTCTTCACTGGATTTTAAGC | 676 | | | | | |
| Db | 635 | AATCTGGTTCTCTTACAGTTCACTTCAACCTTCTTCACTGGATTTTAAGC | 694 | | | | | |
| QY | 677 | AACCTATAAACATCACATCACAGCTCAACAGTGGCAAATTTCGAGACTCTCTGTC | 736 | | | | | Query Match Best Local Similarity Matches |
| Db | 695 | AACCTATAAACATCACATCACAGCTCAACAGTGGCAAATTTCGAGACTCTCTGTC | 754 | | | | | 97.2%; Pred. No. 1.4e-189; 0; Minimatches 22; Indels 0; Gaps 0; |
| QY | 737 | ACGGTGAAGCAGCAGCTTCTCTGATAGAGTCTCACTCTGTCACTGGCTGGAGGCA | 796 | | | | | 75 |
| Db | 755 | ACGGTGAAGCAGCAGCTTCTCTGATAGAGTCTCACTCTGTCACTGGCTGGAGGCA | 814 | | | | | 75 |
| QY | 797 | G 797 | | | | | | 75 |
| Db | 815 | G 815 | | | | | | 75 |
| RESULTS | 6 | AK095314 | | | | | | 197 GCTCAGCTAAATCATGCAACAGGGCCATTCCCTGACCTTGTGACACAG 256 |
| LOCUS | | | 3564 bp | | | | | 215 GCTCAGCTAAATCATGCAACAGGGCCATTCCCTGACCTTGTGACACAG 274 |
| DEFINITION | | Homo sapiens cDNA FLJ37995 | mRNA | | | | | |
| | | | linear | | | | | |
| | | | clone CTONG2011825, moderately similar to CARBONIC ANHYDRASE (EC 4.2.1.1). | | | | | |
| ACCESSION | | | | | | | | 257 AGAACAAATCAGTTCTGGTGTCTCACTGGAAAGTACAGGTACGGAGGTT 316 |

Page 6

| | | | |
|----|-----|---|-----|
| Db | 275 | AGGAAACATCACTAAGTCTGCCTGGGTCTCTCACTGGAAAGCTAACGGTTACGGCGAGTTC | 334 |
| Qy | 317 | ACCTTCACTGGGGTCCGTGATCAACACGGCTCGGACACATAGTAGATGGATGAGCT | 376 |
| Db | 335 | ACCTTCACTGGGGTCCGTGATCAACACGGCTCGGACACATAGTAGATGGATGAGCT | 394 |
| Qy | 377 | ATGCTGAGAGGTCCATGTTCACTGGAAATTGACAAATACCCCAGCTTGTGAGG | 436 |
| Db | 395 | ATGCTGAGAGGTCCATGTTCACTGGAAATTGACAAATACCCCAGCTTGTGAGG | 454 |
| Qy | 437 | CAGCTCATGACCCATGACACTGGCTGCTGGAGTGTCTGGCTGTTTAACAGATTCGACCTA | 496 |
| Db | 455 | CAGCTCATGACCCATGACACTGGCTGCTGGCTGTTTAACAGATTCGACCTA | 514 |
| Qy | 497 | ATTCCAAATCTGAAAAGATAATTACTGACACTTGGATTCCTAAAGAAAAGGTAAACAA | 556 |
| Db | 515 | ATTCCAAATCTGAAAAGATAATTACTGACACTTGGATTCCTAAAGAAAAGGTAAACAA | 574 |
| Qy | 557 | CTCGATTGACAAATTGGACCTATTGCTCTGGTCCACCATCTGGACTACTGGACAT | 616 |
| Db | 575 | CTCGATTGACAAATTGGACCTATTGCTCTGGTCCACCATCTGGACTACTGGACAT | 634 |
| Qy | 617 | ATCCTGGTCTCTAACAGTCCACCTCTCTGGAGGTGCAACGGATGTTTAAGG | 676 |
| Db | 635 | ATCCTGGTCTCTAACAGTCCACCTCTGGAGGTGCAACGGATGTTTAAGG | 694 |
| Qy | 677 | AACCTATAAACATGCGCTCTAACAGGTGCCAACATTGGAGTCTCTGGACAGGG | 736 |
| Db | 695 | AACCTATAAACATGCGCTCTAACAGGTGCCAACATTGGAGTCTCTGGACAGGG | 754 |
| Qy | 737 | AGGGTAACGAGCACCTTCTGATAGATAGAGTCTACTCTGTCACCCAGGTTGGAGGG | 796 |
| Db | 755 | AGGGTAACGAGCACCTTCTGATAGATAGAGTCTACTCTGTCACCCAGGTTGGAGGG | 814 |
| Qy | 797 | G 797 | |
| Qy | 815 | G 815 | |
| Db | | | |

RESULT 7 AX17267 TORCS 100-DEC-2003

| | | | |
|------------|---|---|--|
| DEFINITION | Sequence 15 from patent WO2081517. | | |
| ACCESSION | AX817267 | | |
| VERSION | AX817267.1 | | |
| KEYWORDS | GI:39722654 | | |
| ORGANISM | Homo sapiens Homo sapiens | Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo. | |
| REFERENCE | 1 | Zhong, M., Decistofaro, M.F., Padigaru, M., Miller, C., Tchernev, V., Rastelli, L., Paturra, M., Shimbets, R.A., Shenoy, S., Malynkar, U., Ballinger, R., Gerlach, V., Keruda, R., Guo, X., Wolenc, A., Eisen, A., Vernet, C.A., Li, L., Casman, S., Tailion, B., and Liu, X. | |
| AUTHORS | Zhong, M., Anderson, D., Decistofaro, M.F., Padigaru, M., Miller, C., Tchernev, V., Rastelli, L., Paturra, M., Shimbets, R.A., Shenoy, S., Malynkar, U., Ballinger, R., Gerlach, V., Keruda, R., Guo, X., Zerhusen, B., Andrew, D., Mezes, P., Baumgartner, J., Taubier, R.J., Pena, C., Fernandes, E., Smithson, G., and Liu, X. | | |
| SOURCE | | | |
| JOURNAL | | | |
| TITLE | Novel polypeptides and nucleic acids encoded thereby | | |
| FEATURES | source | Patent: WO 2081517-A-15-17-OCT-2002; Curagen Corporation (US) Location/Qualifiers 1. .828 | |

| | |
|-----------|---|
| REFERENCE | Mammalia: Eutheria; Primates; Catarrhini; Hominoidea; Homo. |
| AUTHORS | Decristofaro, M.F., Padigaru, M., Miller, C., Tchernev, V., Zhong, H., Zheng, M., Anderson, D., Ballinger, R., Gerlach, V., Spytek, K.A., Rastelli, L., Kekuda, R., Guo, X., Zerhusen, B., Andrew, D., Mezes, P., Patrakuram, M., Burgess, C.E., Eisen, A., Wolanc, J., Baumgartner, J., Shimkets, R.A., Gusev, V., Vernet, C.A., Taupier, R.J., Pana, C., Shenoy, S., Li, L., Casman, S., Bolog, F., Fernandes, E., Smithson, G., Malayankar, U., Tailor, B. and Liu, X. |
| TITLE | Novel polypeptides and nucleic acids encoded thereby |
| JOURNAL | Patent : WO 02081517-A1 17 OCT-2002; |
| FEATURES | Curagen Corporation (US) Location/Qualifiers 1. 833 /organism="Homo sapiens" /mol_type="unassigned DNA" /note="unnamed protein product" 21. 809 /codon_start=1 /protein_id="CAE84340.1" /db_xref=GI:3722657 /translation="MSRLSWSYCEFHNGPTEKKEFFPLADGDOQSPPIEKTKFKVYKDSS DHSQSHIVDVSAYAELHVWNSDXYPSPVFEAHAPDGLAVLGVFLQIGPNSQSKQ ITDTLDSIKEKGKQPRTFNFDLSSJLPPSNDYWTWPEGSLTVPPLIESVTWVLVKQDIN ISSQQLAKFRSLLCTAEGERAAFLVSNHRPFQQLGRKVRASH" |
| ORIGIN | Query Match Score 728.8%; DB 6; Length 83; Best Local Similarity 97.1%; Pred. No. 5..e-185; Matches 742; Conservative 0; Mismatches 22; Indels 0; Gaps 0; |
| source | 34 ATGTCGAGGTCACTGGATAACGGTCACTGGTCAATTACTGGAAATT 93 1 AGTCGAGGTCACTGGATAACGGTCACTGGTCAATTACTGGAAATT 80 21 ATGTCGAGGTCACTGGATAACGGTCACTGGTCAATTACTGGAAATT 80 34 ATGTCGAGGTCACTGGATAACGGTCACTGGTCAATTACTGGAAATT 93 94 TTCCCTATTGCTGATGGTCACTGGATAACGGTCACTGGTCAATTACTGGAAATT 153 81 TTCCCTATTGCTGATGGTCACTGGATAACGGTCACTGGTCAATTACTGGAAATT 140 154 ATGACTCTTCTCCGACACTTAGTCACTGGATAACGGTCACTGGTCAATTACTGGAAATT 213 141 ATGACTCTTCTCCGACACTTAGTCACTGGATAACGGTCACTGGTCAATTACTGGAAATT 200 Db |
| Db | 214 AGGAACAGGGCATCTTCACTGGATAACGGTCACTGGTCAATTACTGGAAATT 273 201 AGGACACGGCATTCTTCACTGGATAACGGTCACTGGTCAATTACTGGAAATT 260 274 CCGGGTGTCTCTCACTGGATAACGGTCACTGGTCAATTACTGGAAATT 333 261 CCGGGTGTCTCTCACTGGATAACGGTCACTGGTCAATTACTGGAAATT 320 Db |
| Db | 334 GTGTGATGCCACGCGTCCGGACACATAGTGACGCTGATGGTCACTGGTCAATTACTGGAAATT 393 321 GCGTGTGACCACTGGCTCCGACATAGTGACGCTGATGGTCACTGGTCAATTACTGGAAATT 380 Db |
| Db | 394 GTGTGTCATGGATTCAAGAAATACTGGATAACGGTCACTGGTCAATTACTGGAAATT 453 381 GTGTGTCATGGATTCAAGAAATACTGGATAACGGTCACTGGTCAATTACTGGAAATT 440 Db |
| Db | 454 GGACTGGTGTCTGGACTGTTACAGGTTGAGCTGAACTTAATCTGGAAATT 513 441 GGACTGGCTCTGGACTGTTACAGGTTGAGCTGAACTTAATCTGGAAATT 500 Db |
| Db | 514 ATACTGACAATCTGGATCATAAGAAAGGTTAACAAACTGGTCACTGGAAATT 573 501 ATACTGACAATCTGGATCATAAGAAAGGTTAACAAACTGGTCACTGGAAATT 560 Db |
| Db | 574 GACCTATCTCTGGCTCCACATCCGGACTACTGGACATATCTGGTTCTCTTACA 633 561 GACCTATCTCTGGCTCCACATCCGGACTACTGGACATATCTGGTTCTCTTACA 620 |
| RESULTS | RESULT 9 AF231123 LOCUS AF231123 mRNA linear ROD 20-APR-2001 DEFINITION Mus musculus carbonic anhydrase XIII (Car13) mRNA, complete cds. VERSION AF231123.1 GI:13249092 KEYWORDS Mus musculus (house mouse) ORGANISM Mus musculus Mammalia; Etheria; Rodentia; Sciurogathii; Muridae; Murine; Mus; Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Buteleostomi; REVERSE Hewett-Emmett,D. AUTHORS Chegwidden,W.R.; Carter,N.D. and Edwards,Y.H. (Eds.) ; TITLE Evolution and distribution of the carbonic anhydrase gene families (in) Chegwidden,W.R.; Carter,N.D. and Edwards,Y.H. (Eds.) ; JOURNAL THE CARBONIC ANHYDRASES: NEW HORIZONS: 29-76; PUBLISHER Birkhauser Verlag, Basel (2000) REFERENCE 1 (bases 1 to 1052) 2 (bases 1 to 1052) AUTHORS Hewett-Emmett,D. and Shimmin,L.C. TITLE Characterization and evolution of two new members of the alpha-carbonic anhydrase gene family in mouse: Car13 and Car15 JOURNAL Unpublished REFERENCE 3 (bases 1 to 1052) AUTHORS Shimmin,L.C. and Hewett-Emmett,D. TITLE Direct Submission JOURNAL Submitted (03-FEB-2000) Human Genetics Center, The University of Texas-Houston School of Public Health, P.O. Box 20334, Houston, TX 77225-0334, USA FEATURES Location/Qualifiers source 1. 1052 /organism="Mus musculus" /mol_type="mRNA" /strain="C3H" /db_xref="taxon:10090" /clone="IMAGE: 119823" /cell_line="C2C12" /clone_lib="Barstead mouse myotubes MPL-RBS" Db 1. 1052 /gene="Car13" /product="carbonic anhydrase XIII" /EC_number="4.2.1.1" /note="orthologous to Bq1.3 Homo sapiens genomic sequence in working draft NT_008117.2 and adjacent to CA1, CA3 and CA2 genes; this region is syntenic with mouse chromosome 3 region containing Car1, Car2 and Car3" /cdon_start=1 /cdon_end=1 /protein_id="AAK16672.1" /db_xref="GI:13249093" /translation="MWARLSGTYGEENGPHWNELPFIADGDOQSPPIEKTKFVKYDSSLRPLSIKYDPAKSILTSNSGHSENFDDFTDLSVLRGGPTGNTYRLQFLHWGSDADRHGSRAHVDVQYRANLWTHWNSSPSYVAAESDGLAVNQPLQKQPSITDLDISREKGKQPRFTNPDLCUPLPSNDWYTFGSLTVPPLESVTWVLKQPSISSQQQLARFSLLCTAEGESAFAELLSNRPPQPLKSRVRAFSY" |
| ORIGIN | Query Match Score 590; DB 10; Length 1052; Best Local Similarity 94.2%; Pred. No. 1..e-147; Matches 665; Conservative 0; Mismatches 125; Indels 0; Gaps |

/translation="MARLSWGYGEHNGPITHNIELLPIADGDOOSPIEKTKKEYKYDSS
LRPLSIKYDBASAKLISNGSHSENVDDEDTEDKSVLRGGPLTGNYRLQFLHLINGSD
DHGSERHVDEVRAAELESDDGIAVLEFLQGEHNFQLQK
ITDILDSIKEKGKOTRFTNDFPLCLIPSSWMDWTYPSLUTVPPLLESTWIVLKQPI
ISSQLQARPLSLCTAEGESAFAELLSNRPPQPLKGRRVTRASY"

49 .813
/gene="Car13"
/not="carb_anhydride; Region: Eukaryotic-type carbonic
anhydrase"
/db_xref="CDD:pfam00194"

ORIGIN

| Query | Match | Score | Length | Pred. | No. | Indels | Gaps | 0; |
|---------|-------|--|--------|-------|-----|------------|------|----|
| Qy | 66 | Best Local Similarity | 84.2% | Pred. | No. | 1..1e-147; | | |
| Matches | 65 | Conservative | 84.2% | | | | | |
| Db | 8 | CGGCTCGAGTTCCACCCGGGGACCATGTGAGGGTCAGCTGGGATAACCGGAGCAC | 67 | | | | | |
| Db | 8 | CGTGCACTGGCTTCACTGAAAGGAGCTGGCTGAGCTGGGATACGGCGAGCAC | 67 | | | | | |
| Qy | 68 | ACGGTCCATTACTGAGGGATTTCCTATTGATGATGCTGATCGAACTTCGAA | 127 | | | | | |
| Db | 68 | ACGGTCCATTACTGAACTGGCTTCACTGAAAGGAGCTGGCTGAGCTGGGATACGGCGAGCAC | 127 | | | | | |
| Qy | 128 | TGAGATTAACACRANGAAGTGAATAATGACCTCTCCGACCACTAGTATGAT | 187 | | | | | |
| Db | 128 | TGAGATTAACACRANGAAGTGAATAATGACCTCTCCGACCACTAGTATGAT | 187 | | | | | |
| Qy | 188 | ATGACCAAAGCTCAGCTAAATCATCGAACAAAGCGGCCATTCTTCATGTTGACTTTG | 247 | | | | | |
| Db | 188 | ATGATCCCTGGCTAGCTAAATATCGAACAAATGTCGCACTCTCAACGTTGACTTTG | 247 | | | | | |
| Qy | 248 | ATGACACAGACAAATCAGCTCTGCCTGCTCTCAGCTGAAAGCTACAGTTAC | 307 | | | | | |
| Db | 248 | ACGACACGGAGGAAATCACTGAGCTCTGGCTGGGGTCTCTCATCGGAACTACAGTTGC | 307 | | | | | |
| Qy | 308 | GGAGGGTCACTCTCACTGGGGTCCGGTGTGATGACACGGCTCGAGCACATAGTAGATG | 367 | | | | | |
| Db | 308 | GGGAGTCTCATTGGCACTGGGCTCAGGGTGTGATGACACGGCTCGAGCACATAGTAGATG | 367 | | | | | |
| Qy | 368 | GAGTGTGCTATGTCGAGAGCTCCATGTTGTCACTGGATTCAAGAAATACCCGAGCT | 427 | | | | | |
| Db | 368 | GAGTGTGCTATGTCGAGAGCTCCATGTTGTCACTGGATTCAAGAAATACCCGAGCT | 427 | | | | | |
| Qy | 428 | TGTTGAGGCACTCGAACAGATGGCACTGCACTGAAAGTACTGACACTTGGATTCA | 487 | | | | | |
| Db | 428 | TGTTGAGGCACTCGAACAGATGGCACTGCACTGAAAGTACTGACACTTGGATTCA | 487 | | | | | |
| Qy | 488 | GTGAACTTAATCCCAACTGCAAAAGATTACTGACACTTGGATTCCATGGAAAGG | 547 | | | | | |
| Db | 488 | GGAAACACATCTTCACTGCACTGAAAGTACTGACACTTGGATTCCATGGAAAGG | 547 | | | | | |
| Qy | 548 | GTAACACAAACTGCAAAATTGACCTATTGCTCTGTTGACCTATCTGGACT | 607 | | | | | |
| Db | 548 | GTAACACAAACTGCAAAATTGACCTATTGCTCTGTTGACCTATCTGGACT | 607 | | | | | |
| Qy | 608 | ACTGGGATATCTGGCTCTCTTACAGCTTCAACTCCACCTGACATATTGATGGATTG | 667 | | | | | |
| Db | 608 | ACTGGGATATCTGGCTCTCTGACAGTTCACCTCTGGACT | 667 | | | | | |
| Qy | 668 | TTTAAAGCAACTTAAATACATAGCTCTCAAGCTGGCCAATTTGAGCTCTCTCT | 727 | | | | | |
| Db | 668 | TTTTAAAGCAACTTAAATACATAGCTCTCAAGCTGGCCAATTTGAGCTCTCTCT | 727 | | | | | |
| Qy | 728 | GCACAGGGGAGGTGAGCAAGGAGCTCTGATGAGTCACTCTGTCACCCAGGC | 787 | | | | | |
| Db | 728 | GCACAGGTGAGGAGAGTGCCTGCTTAACTCCAGTCAACAGCTACACAGCC | 787 | | | | | |
| Qy | 788 | TGGAGGGCAG 797 | | | | | | |
| Db | 788 | TGAGGGCC 797 | | | | | | |

RESULT 11 AC103408/C
DEFINITION Rattus norvegicus clone CH230-270112, *** SEQUENCING IN PROGRESS
***, 8 unsorted pieces.
ACCESSION AC109408
VERSION HTGS_008.4 GI: 231.01181
KEYWORD HTGS_DRAFT; HTGS_PHASE1; HTGS_ENRICHED.
ORGANISM Rattus norvegicus
MATERIALS Butheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1. (bases 1 to 267093)
Muzny, D. Marie., Metzker, M. Lee., Abramzon, S., Adams, C., Alder, J.,
Allen, C., Allen, H., Alsbrooks, S., Amir, A., Anguiano, D.,
Ayalelebchi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H.,
Baldwin, D., Bandaranaike, D., Barber, M., Barnstede, M.,
Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M.,
Bryant, N., Buhay, C., Burch, K., Calderon, B.,
Carmena, V., Carter, K., Cavazos, L., Cesiar, H., Center, A.,
Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Z., Chu, J.,
Cleveland, C., Cochrell, R., Cox, C., Coyne, M., Cree, A., D'Souza, L.,
Dayila, M.L., Davis, C., Davy-Carroll, L., De Andra, C., Dederich, D.,
Delgado, O., Denison, S., Deramo, C., Ding, Y., Dinh, H., Divya, K.,
Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Faves, K.,
Egan, A., Escott, M., Eugene, C., Evans, C.A., Evans, T., Fan, G.,
Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P.,
Fraser, C.M., Gabisi, A., Ganta, R., Garcia, R., Garner, T., Garza, M.,
Gebrgeorgis, E., Geer, K., Gall, R., Grady, M., Guerra, W.,
Gunaratne, P., Haaland, W., Hamil, C., Hamilton, K.,
Harvey, Y., Havlak, P., Hawes, A., Hernandez, J.,
Hernandez, R., Hines, S., Hladin, S., Hodgeson, A., Hogues, M.,
Hollins, B., Howell, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A.,
Jackson, L., Jacob, L., Jiang, H., Johnson, R., Jolivet, A.,
Karpathy, S., Kelly, S., Khan, Z., King, L., Kovar, C.,
Kowals, C., Kraft, C.L., Lebow, H., Levan, J.J., Lewis, L., Li, Z., Liu, J.,
Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J.,
Loreschewa, L., Loulseged, H., Lorado, R.J., Lu, X., Ma, J.,
Maheshwari, M., Mahindarne, M., Matmoud, M., Mallory, K., Mangum, A.,
Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E.,
Mawhinney, S., McLeod, M.P., McNeill, T.Z., Meenan, E.,
Milosav-Jevic, A., Popovic, D., Primus, E., Pu, L.-L.,
Puato, M., Quiroz, J., Rachlin, E., Reeves, K., Reger, M.A., Reigh, R.,
Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F.,
Rives, C., Rockey, T., Rojas, A., Rose, M., Ruiz, S.J.,
Sanders, M., Savery, G., Scheer, S., Scott, G., Shatsman, S., Shen, H.,
Pasternak, S., Paul, H., Perez, L., Perez, C.D., Smajs, D.,
Plopper, F., Poindexter, A., Popovic, D., Primus, E., Pu, L.-L.,
Reilly, B., Thomas, N., Thomas, S., Tingay, S., Trejos, Z., Usmani, K.,
Valas, R., Vera, V., Villasana, D., Waldron, J., Walker, B., Wang, J.,
Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F.,
Williams, G., Willson, R., Wiesen, H., Worley, K.,
Wright, D., Wright, R., Wu, J., Yakuib, S., Yen, J., Yoon, L., Yoon, V.,
Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von
Niederhausern, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O.,
Weinstock, G., and Gibbs, R.A.

TITLE Direct Submission
JOURNAL Unpublished
AUTHORS 2 (bases 1 to 267093)
TITLE Direct Submission
JOURNAL Submitted (04 FEB-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 267093)
REFERENCE Rat Genome Sequencing Consortium.

TITLE Direct Submission
JOURNAL Submitted (09-OCT-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

COMMENT On Sep 18, 2002 this sequence version replaced gi:21737978.
 The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole Genome shotgun sequence only contigs will be indicated in the feature table.

Genome Center
 Center code: BCM
 Web site: <http://www.hgsc.bcm.tmc.edu/>
 Contact: hgsc-help@bcm.tmc.edu

Project Information

Center project name: GOBE
 Center clone name: CH230-270I12

Summary Statistics

Assembly program: Phrap, version 0.990329
 Consensus quality: 205376 bases at least Q40
 Consensus quality: 210255 bases at least Q30
 Consensus quality: 213384 bases at least Q20
 Estimated insert size: 233774; sum-of-contigs estimation
 Quality coverage: 4x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length.
 (see http://www.hgsc.bcm.edu/docs/Genbank/draft_data.html).
 * This is a 'working draft' sequence. It currently
 * consists of 8 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

* 1 10796: contig of 10796 bp in length
 * 10797 10896: gap of unknown length
 * 10897 256231: contig of 245335 bp in length
 * 256232 256331: gap of unknown length
 * 256332 257997: contig of 1666 bp in length
 * 257998 258097: gap of unknown length
 * 258098 259380: contig of 1283 bp in length
 * 259381 259480: gap of unknown length
 * 259481 261177: contig of 1697 bp in length
 * 261178 262887: contig of unknown length
 * 262888 262988: gap of unknown length
 * 262988 265407: contig of 2420 bp in length
 * 265408 265503 267093: contig of 1586 bp in length.

FEATURES

source
 Location/Qualifiers
 1. .267093
 /organism="Rattus norvegicus"
 /mol_type="genomic DNA"
 /db_xref="taxon:10116"
 /clone_id="CH230-270I12"
 misc_feature
 1. .10796
 /note="hgsv_end_extension
 clone_end:Sp6"
 misc_feature
 9659. .10796
 /note="hgsv_end_extension
 clone_end:Sp6"
 misc_feature
 9807. .10667
 /note="clone_boundary
 clone_end:Sp6"

| AC112526 | AC112526 | Rattus norvegicus clone CH230-223F17, *** SEQUENCING IN PROGRESS | JOURNAL | COMMENT |
|---|--|--|-----------------|---------|
| DEFINITION | LOCUS | 266620 bp DNA | HTG 13-MAY-2003 | |
| *** 6 unorderded pieces. | | | | |
| ACCESSION | AC112526 | | | |
| VERSION | AC112526_5 | GT:30578729 | | |
| KEYWORDS | HTGS PHASEL | HTGS DRAFT; HTGS_ENRICHED. | | |
| SOURCE | Rattus norvegicus (Norway rat) | | | |
| ORGANISM | Rattus norvegicus | | | |
| Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus. | | | | |
| REFERENCE | 1 (bases 1 to 266620) | | | |
| AUTHORS | Muzny, D. Marie., Merzker, M. Lee., Abramzon, S., Adams, C., Alder, J., Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H., Baldwin, D., Bandaranike, D., Barber, M., Barnsteed, M., Benahmed, F., Bisiwal, K., Blair, J., Blenkinsburg, K., Blyth, P., Brown, M., Cardenas, V., Carter, K., Cavazos, I., Caesar, H., Center, A., Chacko, J., Chavez, D., Chen, Y., Chen, Z., Chu, J., Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L., Davila M.L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D., Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K., Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K., Egan, A., Escott, M., Eugene, C., Evans, C.A., Fallis, T., Fan, G., Fraser, C.M., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M., Gebregeorgis, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W., Guraratne, P., Haaland, W., Hahn, C., Hamilton, C., Harvey, Y., Haylak, P., Hawes, A., Henderson, N., Hernandez, J., Hernandez, R., Hines, S., Hladun, S. L., Hodgson, A., Hogenes, M., Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A., Karpathy, S., Kelly, S., Khan, Z., King, L., Kovar, C., Kowis, C., Kraft, C.L., Lebow, H., Levant, J., Lewis, L., Li, Z., Liu, J., Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J., Lorenshewal, L., Louiseged, H., Lozado, R.J., Lu, X., Ma, J., Maheshwari, M., Mahindarachar, M., Mahmood, M., Mallory, K., Mangum, A., Mangum, B., Mapua, P., Martin, K., Martinez, E., Mawhinney, S., McLeod, M.P., McNeill, T.Z., Meenen, E., Milosavljevic, A., Miner, G., Minja, B., Montenayor, J., Moore, S., Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L., Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Norris, S., Pal, S., Parks, K., Pasternak, O., Okwonou, G., Olarriponsaqon, G., Pal, S., Parks, K., Pasternak, S., Paul, H., Perez, A., Perez-L., Pfannkoch, C., Plopper, F., Poindexter, A., Popovic, D., Primus, E., Pu, L.-L., Puaio, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M.A., Reigh, R., Reilly, B., Reilly, T., Ren, Y., Reuter, M., Richards, S., Riggs, F., Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S.J., Sanders, W., Savery, G., Scherer, S., Scott, G., Shatsman, S., Shen, H., Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C.D., Smajs, D., Smeid, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K., Valas, R., Vera, V., Villasana, D., Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, P., Williams, G., Wilson, R., Wilecy, R., Wooden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von Niederauer, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O., Weinstock, G., and Gibbs, R.A. | | | |
| TITLE | Unpublished | | | |
| JOURNAL | Worley, K.C. | | | |
| REFERENCE | 2 (bases 1 to 266620) | | | |
| AUTHORS | Worley, K.C. | | | |
| TITLE | Direct Submission | | | |
| JOURNAL | Submitted (22-PB-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA | | | |
| REFERENCE | 3 (bases 1 to 266620) | | | |
| AUTHORS | Rat Genome Sequencing Consortium | | | |
| TITLE | Direct Submission | | | |
| FEATURES | | | | |
| source | | | | |
| misc_feature | | | | |
| misc_feature | | | | |
| misc_feature | | | | |

```
/note="wgs end_extension  
clone end:T"
```

| Rattus norvegicus (Norway rat) | | | | | | | | | |
|---|--|--|--|--|--|--|--|--|--|
| SOURCE | Rattus norvegicus | | | | | | | | |
| ORGANISM | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; | | | | | | | | |
| MAMMALIA; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; | | | | | | | | | |
| Rattus | | | | | | | | | |
| | 1 (bases 1 to 276378) | | | | | | | | |
| AUTHORS | Muzny, D. Marie., Metzker, M. Lee., Abramzon, S., Adams, C., Alder, J., Allen, C., Allien, H., Alisbrooks, S., Amin, A., Anguiano, D., Anyalebeche, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H., Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F., Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M., Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E., Cardenas, V., Carter, K., Cavazos, I., Ceasar, H., Center, A., Cesario, J., Chavez, D., Chen, G., Chen, Y., Chen, Z., Cox, J., Cree, A., D'Souza, L., Cleveland, C., Cockrell, R., Cox, G., Coy, J., D'Andrea, C., Dederich, D., Davila, M. L., Davis, C., Davy-Carroll, L., De Andra, C., Delgado, O., Denison, S., Deramo, C., Ding, Y., Dinh, H., Divya, K., Draper, H., Dugan-Rocha, S., Dunn, A., Durval, B., Eaves, K., Egan, A., Escott, M., Eugene, C., Evans, C.A., Flagg, N., Forbes, L., Foster, M., Foster, P., Frasier, C.M., Gabisi, A., Ganta, R.R., Garcia, A., Garner, T., Garza, M., Gebregheorgis, E., Geer, D., Dinh, H., Guerra, W., Guevara, W., Gunaratne, P., Haaland, W., Hamil, C., Hamilton, C., Hamilton, K., Harvey, Y., Havlak, P., Hayes, A., Henderson, N., Hernandez, J., Hernandez, R., Hines, S., Hladun, S.L., Hodges, M., Hollins, B., Howell, S.S., Hullik, S., Hume, J., Idlebird, D., Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A., Karpowich, S., Kelly, S., Khan, Z., King, L., Kovari, C., Kowis, C., Kraft, C.L., Lebow, H., Lewis, L., Li, Z., Liu, J., Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J., Loresuhera, L., Louisedrine, H., Lozada, R.J., Ma, J., Mahmood, M., Mahindarne, M., Mangum, A., Mangum, B., Mapta, P., Martin, K., Martin, R., Martinez, E., Mawhinney, S., McLeod, M.P., McNeilly, T.Z., Meener, B., Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S., Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L., Nankervis, C., Neal, D., Newton, R., Nguyen, N., Norris, S., Nwakelomeh, O., Okwonou, G., Olarnpusangon, A., Pal, S., Parks, K., Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkoch, C., Plopper, F., Pointdexter, B., Popovic, D., Primus, B., Pu, L.-L., Puazo, M., Quiroz, J., Rachlin, B., Reeves, K., Regier, M.A., Reigh, R., Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F., Sanders, W., Savery, G., Scherer, S., Scott, G., Shatsman, S., Shen, H., Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C.D., Smajs, D., Sneed, A., Sodergren, B., Sorelle, R., Sosa, J., Steimle, M., Strong, R., Sutton, A., Svatek, A., Taylor, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Uismani, K., Valas, R., Vera, V., Villasana, D., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F., Williams, G., Wilson, R., Wileczky, R., Woodward, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von Niederhausern, A., Weiss, R., Smith, D.R., Holt, R.A., Weinstock, G., and Gibbs, R.A. | | | | | | | | |
| REFERENCE | 1 (bases 1 to 276378) | | | | | | | | |
| AUTHORS | Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E., Cardenas, V., Carter, K., Cavazos, I., Ceasar, H., Center, A., Cesario, J., Chavez, D., Chen, G., Chen, Y., Chen, Z., Cox, J., Cree, A., D'Souza, L., Cleveland, C., Cockrell, R., Cox, G., Coy, J., D'Andrea, C., Dederich, D., Davila, M. L., Davis, C., Davy-Carroll, L., De Andra, C., Delgado, O., Denison, S., Deramo, C., Ding, Y., Dinh, H., Divya, K., Draper, H., Dugan-Rocha, S., Dunn, A., Durval, B., Eaves, K., Egan, A., Escott, M., Eugene, C., Evans, C.A., Flagg, N., Forbes, L., Foster, M., Foster, P., Frasier, C.M., Gabisi, A., Ganta, R.R., Garcia, A., Garner, T., Garza, M., Gebregheorgis, E., Geer, D., Dinh, H., Guerra, W., Guevara, W., Gunaratne, P., Haaland, W., Hamil, C., Hamilton, C., Hamilton, K., Harvey, Y., Havlak, P., Hayes, A., Henderson, N., Hernandez, J., Hernandez, R., Hines, S., Hladun, S.L., Hodges, M., Hollins, B., Howell, S.S., Hullik, S., Hume, J., Idlebird, D., Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A., Karpowich, S., Kelly, S., Khan, Z., King, L., Kovari, C., Kowis, C., Kraft, C.L., Lebow, H., Lewis, L., Li, Z., Liu, J., Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J., Loresuhera, L., Louisedrine, H., Lozada, R.J., Ma, J., Mahmood, M., Mahindarne, M., Mangum, A., Mangum, B., Mapta, P., Martin, K., Martin, R., Martinez, E., Mawhinney, S., McLeod, M.P., McNeilly, T.Z., Meener, B., Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S., Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L., Nankervis, C., Neal, D., Newton, R., Nguyen, N., Norris, S., Nwakelomeh, O., Okwonou, G., Olarnpusangon, A., Pal, S., Parks, K., Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkoch, C., Plopper, F., Pointdexter, B., Popovic, D., Primus, B., Pu, L.-L., Puazo, M., Quiroz, J., Rachlin, B., Reeves, K., Regier, M.A., Reigh, R., Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F., Sanders, W., Savery, G., Scherer, S., Scott, G., Shatsman, S., Shen, H., Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C.D., Smajs, D., Sneed, A., Sodergren, B., Sorelle, R., Sosa, J., Steimle, M., Strong, R., Sutton, A., Svatek, A., Taylor, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Uismani, K., Valas, R., Vera, V., Villasana, D., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F., Williams, G., Wilson, R., Wileczky, R., Woodward, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von Niederhausern, A., Weiss, R., Smith, D.R., Holt, R.A., Weinstock, G., and Gibbs, R.A. | | | | | | | | |
| JOURNAL | Unpublished | | | | | | | | |
| COMMENT | 3 (bases 1 to 276378) | | | | | | | | |
| REFERENCE | 3 (bases 1 to 276378) | | | | | | | | |
| AUTHORS | Worley, K.C. | | | | | | | | |
| JOURNAL | Direct Submission | | | | | | | | |
| COMMENT | The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described | | | | | | | | |
| REFERENCE | (15-NOV-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA | | | | | | | | |
| AUTHORS | On Nov 6, 2002 this sequence version replaced gi:23269974. | | | | | | | | |
| JOURNAL | Direct Submission | | | | | | | | |
| COMMENT | Submitted (116-FEB-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA | | | | | | | | |
| REFERENCE | (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described | | | | | | | | |
| AUTHORS | Submitted (15-NOV-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA | | | | | | | | |
| JOURNAL | Direct Submission | | | | | | | | |
| COMMENT | The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described | | | | | | | | |
| REFERENCE | (15-NOV-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA | | | | | | | | |
| AUTHORS | Submitted (116-FEB-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA | | | | | | | | |
| JOURNAL | Direct Submission | | | | | | | | |
| COMMENT | The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described | | | | | | | | |
| REFERENCE | (15-NOV-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA | | | | | | | | |
| AUTHORS | Submitted (116-FEB-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA | | | | | | | | |
| JOURNAL | Direct Submission | | | | | | | | |
| COMMENT | The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described | | | | | | | | |
| REFERENCE | (15-NOV-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA | | | | | | | | |
| AUTHORS | Submitted (116-FEB-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA | | | | | | | | |
| JOURNAL | Direct Submission | | | | | | | | |
| COMMENT | The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described | | | | | | | | |
| REFERENCE | (15-NOV-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA | | | | | | | | |
| AUTHORS | Submitted (116-FEB-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA | | | | | | | | |
| JOURNAL | Direct Submission | | | | | | | | |
| COMMENT | The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described | | | | | | | | |
| REFERENCE | (15-NOV-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA | | | | | | | | |
| AUTHORS | Submitted (116-FEB-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA | | | | | | | | |
| JOURNAL | Direct Submission | | | | | | | | |
| COMMENT | The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described | | | | | | | | |
| REFERENCE | (15-NOV-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA | | | | | | | | |
| AUTHORS | Submitted (116-FEB-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA | | | | | | | | |
| JOURNAL | Direct Submission | | | | | | | | |
| COMMENT | The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described | | | | | | | | |
| REFERENCE | (15-NOV-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA | | | | | | | | |
| AUTHORS | Submitted (116-FEB-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA | | | | | | | | |
| JOURNAL | Direct Submission | | | | | | | | |
| COMMENT | The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described | | | | | | | | |
| REFERENCE | (15-NOV-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA | | | | | | | | |
| AUTHORS | Submitted (116-FEB-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA | | | | | | | | |
| JOURNAL | Direct Submission | | | | | | | | |
| COMMENT | The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described | | | | | | | | |
| REFERENCE | (15-NOV-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA | | | | | | | | |
| AUTHORS | Submitted (116-FEB-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA | | | | | | | | |
| JOURNAL | Direct Submission | | | | | | | | |
| COMMENT | The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described | | | | | | | | |
| REFERENCE | (15-NOV-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA | | | | | | | | |
| AUTHORS | Submitted (116-FEB-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA | | | | | | | | |
| JOURNAL | Direct Submission | | | | | | | | |
| COMMENT | The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described | | | | | | | | |
| REFERENCE | (15-NOV-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA | | | | | | | | |
| AUTHORS | Submitted (116-FEB-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA | | | | | | | | |
| JOURNAL | Direct Submission | | | | | | | | |
| COMMENT | The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described | | | | | | | | |
| REFERENCE | (15-NOV-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA | | | | | | | | |
| AUTHORS | Submitted (116-FEB-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA | | | | | | | | |
| JOURNAL | Direct Submission | | | | | | | | |
| COMMENT | The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described | | | | | | | | |
| REFERENCE | (15-NOV-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA | | | | | | | | |
| AUTHORS | Submitted (116-FEB-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA | | | | | | | | |
| JOURNAL | Direct Submission | | | | | | | | |
| COMMENT | The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described | | | | | | | | |
| REFERENCE | (15-NOV-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA | | | | | | | | |
| AUTHORS | Submitted (116-FEB-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA | | | | | | | | |
| JOURNAL | Direct Submission | | | | | | | | |
| COMMENT | The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described | | | | | | | | |
| REFERENCE | (15-NOV-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA | | | | | | | | |
| AUTHORS | Submitted (116-FEB-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA | | | | | | | | |
| JOURNAL | Direct Submission | | | | | | | | |
| COMMENT | The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described | | | | | | | | |

Page 13

in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

| FEATURES | source | Query Match | Score 493.2; | DB 6; | Length 688; |
|------------|---|--|--------------|---------------------|-------------|
| FH | Key | Best Local Similarity | 95.7%; | Pred. No. 1.2e-121; | |
| LOCATIONS | Matches | Conservative | 0; | Mismatches | Indels 0; |
| QUALIFIERS | /db_xref="Taxon:9606" | | | | |
| ORIGIN | | | | | |
| QY | 268 | GTTCTGCCTGGTCTCCCTCACTGAAAGCTACAGCTTACGGCAGCTTCACCTTCACCTGG | 3.27 | | |
| Db | 43 | GTTCTGCCTGGTCTCCCTCACTGAAAGCTACAGCTTACGGCAGCTTCACCTTCACCTGG | 1.02 | | |
| Qy | 328 | GGTCTCCCTGATGAC CACGGCTCGAGCACATAGTAGATGCTGAGCTGAGCTATGCTGAGAG | 3.87 | | |
| Db | 103 | GGTCTCCGTGATGAC CACGGCTCGAGCACATAGTAGATGCTGAGCTATGCTGAGAG | 1.62 | | |
| Qy | 388 | CCTCATGTTGTTACTGAAATTGAGAACATACCCAGCTTGAGGAGCTCATGAA | 4.47 | | |
| Db | 163 | CCTCATGTTGTTACTGAAATTGAGAACATACCCAGCTTGAGGAGCTCATGAA | 2.22 | | |
| Qy | 448 | CCAGATGGACTGGCTGTCCTGGACTGTTACAGATTGCTGAACTTAATCCAACTG | 5.07 | | |
| Db | 223 | CCAGATGGACTGGCTGTCCTGGACTGTTACAGATTGCTGAACTTAATCCAACTG | 2.82 | | |
| Qy | 508 | CAAAGATTACTGACACTTGGATTCCATTAAAGAAAAGGTAACAAACACTGATTCA | 5.67 | | |
| Db | 283 | CAAAGATTACTGACACTTGGATTCCATTAAAGAAAAGGCTAACAAACACTGATTCA | 3.42 | | |
| Qy | 568 | AATTTCGACCTATTGCTCTGTTACCATTCCTGGACTGACTGGACATATCCTGGTCT | 6.27 | | |
| Db | 343 | AATTTCGACCTATTGCTCTGTTACCATTCCTGGACTGACTGGACATATCCTGGTCT | 4.02 | | |
| Qy | 628 | CTTACAGTCCACCTCTTGGAGGTGACATGGTATTGTTAAAGCAACCTTAAC | 6.87 | | |
| Db | 403 | CTTACAGTCCACCTCTTGGAGGTGACATGGTATTGTTAAAGCAACCTTAAC | 4.62 | | |
| Qy | 688 | ATCAGCTCTAACAGCTGGCCAATTTCGAGTCCTGTCACAGCGGGTGNAGCA | 7.47 | | |
| Db | 463 | ATCAGCTCTAACAGCTGGCCAATTTCGAGTCCTGTCACAGCGGGTGNAGCA | 5.22 | | |
| Qy | 748 | GGAGCTTCTCTGATAGAGTCTCAGCTCTGTCACCCAGGTGGCAG | 7.97 | | |
| Db | 523 | GGAGCTTCTCTGAGGATCACCGGCCAACACCAGCCCTTAAGGGCCG | 5.72 | | |
| RESULT | 15 | | | | |
| LOCUS | AK093753 | AK093753 | 2403 bp | mRNA | linear |
| DEFINITION | Homo sapiens cDNA FLJ36434 fis | | | PRI 06-SEP-2002. | |
| ACCESSION | AK093753 | | | | |
| VERSION | AK093753.1 | GI:21752680 | | | |
| KEYWORDS | oligo capping; fis (full insert sequence). | | | | |
| SOURCE | Homo sapiens (human) | | | | |
| ORGANISM | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Butheria; Primates; Catarrini; Hominidae; Homo. | | | | |
| REFERENCE | 1 | Ishibashi, T.; Kaneko, K.; Yosida, M.; Watanabe, S.; Ishida, S., Ono, Y.; Horita, T.; Hirada, S.; Murakawa, K.; Takiguchi, S.; Kusano, J.; Watanabe, M.; Fujimori, K.; Tanai, H.; Ishida, M., Yamashita, H.; Chiba, Y.; Sugiyama, T.; Irie, R.; Sato, H., Ota, T.; Wakamatsu, A.; Ishii, S.; Yamamoto, J.; Isono, Y.; Kawai, Hio, Y.; Saito, K.; Nishikawa, T.; Kimura, K.; Matsumo, K.; Nakamura, Y.; Sekine, M.; Kikuchi, H.; Kanda, K.; Nagatsuma, M., Takahashi, T.; Fujii, A.; Oshima, A.; Sugiyama, A.; Nagai, K.; and Isocai, T.; Sugano, S.; Nagahashi, K.; Masuho, Y.; Kawamura, B.; Suzuki, Y. | | | |

| | | | |
|----|------|---|------|
| Db | 1226 | GGATCCATTAAAGAAAAAGGTTAACAAACTGATTCAACAAATTGTACCTATGTCCTC | 1285 |
| Qy | 588 | GCTTCCACCATCTGGAACTACTGGACATATCGTCTCTTACAGTTCCACCTTC | 647 |
| Db | 1286 | GCTTCCACCATCTGGAACTACTGGACATATCGTCTCTTACAGTTCCACCTTC | 1345 |
| Qy | 648 | TGAGAGTGACATGGATTGGTTAAAGCAACCTATAAACATCAGCTCAACAGT | 704 |
| Db | 1346 | TGAGAGTGACATGGATTGGTTAAAGCAACCTATAAACATCAGCTCAACAGT | 1402 |

Search completed: September 13, 2004, 23:48:37
Job time : 3648 secs

GenCore version 5.1.6
 Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 13, 2004, 22:17:56 ; Search time 2472 Seconds
 (without alignments) 11005.035 Million cell updates/sec

Title: US-10-069-434-4
 Perfect score: 911
 Sequence: 1 cggaaatccggctcgaggccc.....ccccaaatgcgcggggaaaa 911

Scoring table: IDENTITY_NUC
 Gapop 10_0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0

Maximum DB seq length: 20000000000

Post-processing: Maximum Match 0%
 Listing First 45 summaries

Database : EST:
 1: em_estba:
 2: em_esrhum:
 3: em_estin:
 4: em_estmu:
 5: em_estov:
 6: em_estpl:
 7: em_estro:
 8: em_htc:
 9: gb_est1:
 10: gb_est2:
 11: gb_htc:
 12: gb_est3:
 13: gb_est4:
 14: gb_est5:
 15: em_estfun:
 16: em_estom:
 17: em_gss_hum:
 18: em_gss_inx:
 19: em_gss_ipn:
 20: em_gss_vrt:
 21: em_gss_fun:
 22: em_gss_man:
 23: em_gss_mus:
 24: em_gss_pro:
 25: em_gss_rnd:
 26: em_gss_phg:
 27: em_gss_vrl:
 28: gb_gss1:
 29: gb_gss2:
 *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match Length | DB ID | Description |
|------------|-------|--------------------|----------|---|
| 1 | 728.8 | 80.0 | AY409020 | AY409022 Mus musculus CB317335 AGENCOURT BY924401 |
| 2 | 725.4 | 79.3 | AY409021 | BY44350 BY744350 BY743788 BY744810 BY744678 BY744678 BY744678 CF768871 CEB000987 BY722663 BY722663 BE237415 1467201 MA BU169155 AGENCOURT BU667406 v04d09_r BU357489 603407609 AA726331 v039h08_r CRF288079 AGENCOURT BX674347 BX674347 AA537707 vK44h09_r AK030666 Mus musculus BI764023 60304224 BU523425 AGENCOURT BM427234 Pkf2n_pk0 BU523065 AGENCOURT BQ954504 AGENCOURT BU523833 AGENCOURT BU657013 C119ad6_z BQ922926 AGENCOURT BU656112 C106f05_z BQ928426 AGENCOURT AR033719 Mus musculus BI838466 603083447 AA645297 v881a11_r BU661158 C170h12_z BQ960607 AGENCOURT BU660624 C162d03_z BU408753 603159466 BU412622 603159627 BU8595698 BB859598 |
| 3 | 675.6 | 67.6 | AK010166 | AY409020 Homo sapiens Hominidae; Catarrhini; Vertebrata; Craniata; Euteleostomi; Mammalia; Butheria; Primates; Catarrhini; Hominidae; Homo. Clark,A.G., Glangowski,S., Nielson,R., Thomas,P., Kejariwal,A., Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B., Ferreria,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J., Adams,M.D., and Caigill,M. |
| 4 | 64.8 | 64.8 | AK010166 | Intererring nonneutral evolution from human-chimp-mouse orthologous gene trios: Science 302 (5652), 1960-1963 (2003) 1 (bases 1 to 789) Clark,A.G., Glangowski,S., Nielson,R., Thomas,P., Kejariwal,A., Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B., Ferreria,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J., Adams,M.D., and Caigill,M. Direct Submission |

ALIGNMENTS

| RESULT | LOCUS | DEFINITION | ACCESSION | VERSION | KEYWORDS | SOURCE | ORGANISM | JOURNAL | REFERENCE |
|--------|----------|---|-----------|------------|-------------|--------|----------------------|--------------------------------------|---|
| 1 | AY409020 | Homo sapiens HCM3418 gene, VIRTUAL TRANSCRIPT, partial sequence, Genomic survey sequence. | AY409020 | AY109020.1 | GT:39764988 | GSS. | Homo sapiens (human) | Science 302 (5652), 1960-1963 (2003) | Clark,A.G., Glangowski,S., Nielson,R., Thomas,P., Kejariwal,A., Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B., Ferreria,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J., Adams,M.D., and Caigill,M. |

| | | | |
|----------|--|-------------------------|--|
| JOURNAL | Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive, Rockville, MD 20850, USA | DEFINITION | Pan troglodytes HCM3418 gene, VIRTUAL TRANSCRIPT, partial sequence. |
| COMMENT | This sequence was made by sequencing genomic exons and ordering them based on alignment. | GENOMIC SURVEY SEQUENCE | AY409021 |
| FEATURES | Location/Qualifiers | KEYWORDS | GR:39764989 |
| source | 1..789 /organism="Homo sapiens" /mol_type="genomic DNA" <1..>789 /locus_tag="HCM3418" | SOURCE | GSS |
| ORIGIN | Query Match Score 728.8; Pred. No. 2.1e-200; Best Local Similarity 97.1%; Matches 742; Conservative 0; Mismatches 22; Indels 0; Gaps 0; | ORGANISM | Pan troglodytes (chimpanzee) |
| gene | 34 ATCTCGAAGCTCACTGGGATACCGCAGACAACGGTCCATTCACTGGAGGAAATT 93 1 ATGTCGAGGCTCAGCTGGGATACCGCAGACAACGGTCCATTCACTGGAGGAAATT 60 | REFERENCE | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Buteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan. |
| | 94 TTCCCTATTTGCTATGGTATCAGCAATTCCTCAATGGATTAACCAAAGAGTGAAT 153 61 TTCCCTATTTGCTATGGTATCAGCAATTCCTCAATGGATTAACCAAAGAGTGAAT 120 | AUTHORS | 1 (bases 1 to 789) Clark, A.G., Gianiowski, S., Nielson, R., Thomas, P., Kejariwal, A., Todd, M.A., Tanenbaum, D.M., Civello, D.R., Lu, F., Murphy, B., Ferreria, S., Wang, G., Zheng, X.H., White, T.J., Shinsky, J.J., Adams, M.D., and Cargill, M. |
| ORIGIN | 154 TATGACTCTTCCCTCGACCACCACTTGTCAATGTCAGCTGCTAAATATCTC 213 121 TATGACTCTTCCCTCGACCACCACTTGTCAATGTCAGCTGCTAAATATCTC 180 | TITLE | Infering nonneutral evolution from human-chimp-mouse orthologous gene trios |
| | 214 AGCAACAGGCCATTCTCACTGGAACTTACCTTGACTTGTGACACAGAAATCAGTTCTG 273 181 AGCAACAGGCCATTCTCACTGGAACTTACCTTGACTTGTGACACAGAAATCAGTTCTG 240 | JOURNAL | Science 302 (5652), 1960-1963 (2003) |
| ORIGIN | 274 CGTGGTGGCTCTCTCACTGGAACTTACCTTGACTTGTGACACAGAAATCAGTTCTG 333 241 CGTGGTGGCTCTCTCACTGGAACTTACCTTGACTTGTGACACAGAAATCAGTTCTG 300 | COMMENT | This sequence was made by sequencing genomic exons and ordering them based on alignment. |
| | 34 GCTGATGACCGGACTCCGAGCACATACTGGAACTTACCTTGACTTGTGACACAGAAATCAGTTCTG 393 301 GCTGATGACCGGCTCCAGCAATAAGTAGATGGTGTGAGCTCAT 360 | FEATURES | Gene |
| ORIGIN | 394 GTTGTTCACTGGAAATTCTGACAAATACTCCAGGTTGGTGTGAGCTCATGAAACGAT 453 361 GTTGTTCACTGGAAATTCTGACAAATACTCCAGGTTGGTGTGAGCTCATGAAACGAT 420 | SOURCE | 1 ATGTCGAGGCTCAGCTGGGATACCGCAGACAACGGTCCATTCACTGGAGGAAATT 93 1 ATGTCGAGGCTCAGCTGGGATACCGCAGACAACGGTCCATTCACTGGAGGAAATT 60 |
| | 454 GGACTGGCTCTGGGATCTGGTTTACAGATGGTAACAAACTGGTCAAAATT 513 421 GGACTGGCTCTGGGATCTGGATTTACAGATGGTAACAAACTGGTCAAAATT 480 | ORIGIN | 1. .>789 /locus_tag="HCM3418" |
| ORIGIN | 514 ATTACTGACACTTGTGATTCCTAAAGAAAAGGTAACAAACTGGTCAAAATT 573 481 ATTACTGACACTTGTGATTCCTAAAGAAAAGGTAACAACTGGTCAAAATT 540 | QUERY | Query Match Score 722.4; DB 29; Length 789; Best Local Similarity 96.6%; Pred. No. 1.5e-198; Mismatches 0; Indels 0; Gaps 0; |
| | 574 GACCTATGCTCGCTGCCACCATCTGGGACTACTGGACATACCTGGTTCTCTTACA 633 541 GACCTATGCTCGCTGCCACCATCTGGGACTACTGGACATACCTGGTTCTCTTACA 600 | DB | 1 ATGTCGAGGCTCAGCTGGGATACCGCAGACAACGGTCCATTCACTGGAGGAAATT 93 1 ATGTCGAGGCTCAGCTGGGATACCGCAGACAACGGTCCATTCACTGGAGGAAATT 60 |
| ORIGIN | 634 GTTCACACTCTCTCTGTGAGGTGATCATGATGGTAACTGGAACTTAAACATCGC 693 601 GTTCCACCTCTCTGTGAGGTGATCATGATGGTAACTGGAACTTAAACATCGC 660 | DB | 1 ATGTCGAGGCTCAGCTGGGATACCGCAGACAACGGTCCATTCACTGGAGGAAATT 93 1 ATGTCGAGGCTCAGCTGGGATACCGCAGACAACGGTCCATTCACTGGAGGAAATT 60 |
| | 694 TCTGACAGTGGCAAATTTCGAGTCTCATGATGGTAAAGCAGCACT 753 661 TCTGACAGTGGCAAATTTCGAGTCTCATGATGGTAAAGCAGCACT 720 | DB | 1 ATGTCGAGGCTCAGCTGGGATACCGCAGACAACGGTCCATTCACTGGAGGAAATT 93 1 ATGTCGAGGCTCAGCTGGGATACCGCAGACAACGGTCCATTCACTGGAGGAAATT 60 |
| ORIGIN | 754 TTTCCTGTGATAGAGACTCTCACTCTGTGACCCAGGTGGGGCAG 797 721 TTTCGGTGGCAATCACCCGCCAACACAGCCTAAAGGCCG 764 | DB | 1 ATGTCGAGGCTCAGCTGGGATACCGCAGACAACGGTCCATTCACTGGAGGAAATT 93 1 ATGTCGAGGCTCAGCTGGGATACCGCAGACAACGGTCCATTCACTGGAGGAAATT 60 |
| RESULT | 2 | QY | 454 GGACATGGCTCTCTGGAGGTGTTTACAGATGGTGAACCTTAACTGGAACTTAAACATCGC 513 |
| LOCUS | AY409021 | AY409021 | 421 GGACATGGCTCTCTGGAGGTGTTTACAGATGGTGAACCTTAACTGGAACTTAAACATCGC 480 |

Qy 514 ATTACTGACACTTGGATTCCATTAAAGAAAACGGTAAACAAACTCGATCACAAATT 573
 Db 481 ATTACTGACACTTGGATTCCATTAAAGGTAAACAACTCGATCACAAATT 540

Qy 574 GACCTATGTCCTGCTGCCTCACCATTCTGGATACTTGACATATCCTGGTCTCTTACA 633
 Db 541 GACCCATGTCCTGCTGCCTCACCATTCTGGACTACTGGACATATCCTGGTCTCTTACA 600

Qy 634 GTTCCACCTCTTGTGAGTGTCACTGGATGTTTAAGAACCTTANACATAGC 693
 Db 601 GTTCCACCTCTTGTGAGTGTCACTGGATGTTTAAGAACCTTANACATAGC 660

Qy 694 TCTCAAACAGCTGCCAAATTTCAGTCCTGCAAGCGGAGGTGAACGAGGCT 753
 Db 661 TCTCAAACAGCTGCCAAATTTCAGTCCTGCAAGCGGAGGTGAACGAGGCT 720

Qy 754 TTCTGTGATAGTAGTCTCATCTGTCACTGGACCTGGAGGGCTGGAGGCAG 797
 Db 721 TTCTGTGAGGATATGACCTGGACCTGGAGGGCTGGAGGCAG 764

ORIGIN

Query Match Score 615.1; DB 12; Length 620;
 Best Local Similarity 99.7%; Pred. No. 1.8e-167;
 Matches 616; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 15 AGTCACCCCCAGGGACCATATGCAAGGCTAGCTGGATACCGGAGCAACGCTCC 74
 Db 3 ACTCACCCCCAGGGACCATATGCAAGGCTAGCTGGATACCGGAGCAACGCTCC 62

Qy 75 TATTCACTGGAGGAATTTCCTCATTTGCTGATGGTGTACGCAATCTCCAAATTGAGAT 134
 Db 63 TATTCACTGGAGGAATTTCCTCATTTGCTGATGGTGTACGCAATCTCCAAATTGAGAT 122

Db 135 TAAACCAAAAGGTGAATATGACTTCCTCAATGTTGACTCTTGATGACAC 194
 Db 123 TAAACCAAAAGGTGAATATGACTTCCTCAATGTTGACTCTTGATGACAC 182

Qy 195 AAGCTCAGCTAAATCATCAGAACAGGGCAATTCTTCATGTTGACTCTTGATGACAC 254
 Db 183 AAGCTCAGCTAAATCATCAGAACAGGGCAATTCTTCATGTTGACTCTTGATGACAC 242

Qy 255 AGAGAACAAATCAGTCAGTCAGTCAGGTTACGGCAGGT 314
 Db 243 AGAGAACAAATCAGTCAGTCAGGTTACGGCAGGT 302

Qy 315 TGAACCTCACTGGGTCGCTGATACACAGTACATGTTGAGTGAG 374
 Db 303 TCACTTCACCTGGGTCGGCTGATGACACAGGCTACAGTTACGGCAGGT 362

Qy 375 CTATGCTGAGAGCTCCATGGGATTCAGACAAATCCCCAGCTTGTGAG 434
 Db 363 CTATGCTGAGAGCTCCATGGGATTCAGACAAATCCCCAGCTTGTGAG 422

Qy 435 GGAGCTCTCATGACCAAGATGGACTGGCTCTGGAGTTACAGATTGGTAACC 494
 Db 423 GGCAGCTCTCATGACCAAGATGGACTGGCTCTGGAGTTACAGATTGGTAACC 482

Qy 495 TAATTCGCACTGAAAGATACTGACACTGGCTTGTGTTACTGGATTAGACAAATCCAGTTGTGAG 554
 Db 483 TAATTCCAACTGCAAAAGATACTGACACTGGCTTGTGTTACTGGATTAGACAAATCCAGTTGTGAG 542

COMMENT

DEFINITION mRNA linear EST 04-MAR-2002

VERSION S14K402-2-F11 5 , mRNA

KEYWORDS Sequence.

ORGANISM Homo sapiens (human)

REFERENCE 1 (bases 1 to 620) Mammalia; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.

AUTHORS Kim, N.S., Hahn, Y., Oh, J.H., Lee, J.Y., Ahn, H.Y., Chu, M.Y., Kim, M.R., Oh, K.J., Cheong, J.E., Sohn, H.Y., Kim, J.M., Park, H.S., Kim, S. and Kim, Y.S.

TITLE 21C Frontier Korean EST Project 2001

JOURNAL Unpublished (2002)

CONTACT: Kim, Y.S.

Korea Research Institute of Bioscience & Biotechnology

52 Boeun-dong Yuseong-gu, Daejeon 305-333, South Korea

Fax: +82-42-860-4470

Email: yongsung@mail.kribb.re.kr

Plate: 2 row: F column: 11

High quality sequence stop: 620.

Location/Qualifiers

1..620

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="S14K402-2-F11"

/cell_line="S14K402"

/lab_host="Top10F"

/clone_lab="S14K402"

/note="Organ: Stomach; Vector: pTZ18RPL; Site_1: EcoRI; Site_2: NotI; The poly (A)+ RNA was dephosphorylated with bacterial alkaline phosphatase (BAP) and then decapped with tobacco acid pyrophosphatase (TAP). The decapped intact mRNA was ligated with DNA-RNA linker including Eco RI site by treatment of T4 RNA ligase and the first strand cDNA was synthesized from oligo dT selected mRNA by priming with dT-tailed vector. The dT-tailed vector was adjusted to have about 60nt. The cDNA vector was circularized with E. coli DNA ligase after digestion of EcoRI which site is also included in vector. An RNA strand converted to a DNA strand by Okayama-Berg method. The obtained cDNA vectors were used for transformation of competent cells E. coli Top10F by electroporation method. The cDNA libraries constructed by this method are full-length enriched cDNA library."

RESULT 4

ACCESSION AK010166

LOCUS AK010166

DEFINITION Mus musculus adult male tongue cDNA, RIKEN full-length enriched library. Clone:2310075c21 product:carboxic anhydride 13 , full insert sequence.

VERSION AK010166

KEYWORDS HTC ; CAP trapper.

SOURCE Mus musculus (house mouse)

ORGANISM Mus musculus

ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE Carninci, P. and Hayashizaki, Y.

AUTHORS Carninci, P. and Hayashizaki, Y.

TITLE High-efficiency full-length cDNA cloning

JOURNAL Meth. Enzymol. 303, 19-44 (1999)

MEDLINE 99279253

PUBMED 10496356

REFERENCE Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.

AUTHORS Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.

| | |
|---|---|
| PCR. This library was enriched for full-length clones and was constructed by Clontech Laboratories (Palo Alto, CA). Note: this is a NIH_MGC Library." | Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. |
| ORIGIN | |
| Query Match 60.1%; Best Local Similarity 86.1%; Matches 630; Conservative 0; Mismatches 100; Indels 2; Gaps 2; | Score 548; DB 14; Length 751; |
| Qy 29 GGACCATCTCAGGCTCACTGGGATAACGGTCCATTACTGGAGG 88 | Db 20 GGACCATGCGAGCTGGCTGAGCTGGGTACGGTAGCTGGATG 79 |
| Qy 89 AATTTCCTTAATGCTGATGCTCAATCTCCATTGAGATAAAACAAAG 148 | Db 80 AATTGTTCTCATGCTGATGGTATGGTATGGTATGGTATGGTAAACAAAG 139 |
| Qy 149 TGAATATGACTCTCCGACCACTTAGPATCAAGTATGACCCAGCTCGCTAAA 208 | Db 140 TGAATAGCTCTCACTCCGACCTCTCAACTATGATCCCTCGCTAA 199 |
| Qy 209 TCTACGGACACGGCATCCCTCACTGGTACCTTGACTTGTACAG 268 | Db 200 TCTACGGAAATGGCATTCTCACTGTCAGTCACTGTCAGTCA 259 |
| Qy 269 TTCTGGCTGGTGTCTCTACGGTACAGGTTACCTTCACTGG 328 | Db 260 TTCTGGGGAGGCTCTACGGTACAGGTTGGAGTCACTGG 319 |
| Qy 329 GGTCCGCTGATGACCACCGCTCCGAGCACATAGTAGTGAGCTATGCTGAGAC 388 | Db 320 GGTCAAGGAGATGATCATGGCTCAAGCTGAGTGTGAGTGCAGGC 379 |
| Qy 389 TCTATGTTCTGCTGGATTGACAAATAACCCAGTTGTGAGCAGCTAAC 448 | Db 380 TCTATGTTCTGCTGGATTGACAAATAACCCAGTTGTGAGCAGCTAAC 439 |
| Qy 449 CAGATGGACTGGCTGCTGGTCTGGGACTTTACAGATTTGGAACT 508 | Db 440 CCATGGGTGGTGTCTGGGATATTCTGAGATGGGAAACATCTCACTGC 499 |
| Qy 509 AAAGATTAATGACACTTGGATTCATTAAGAAAGGGTAACACTGGATCACAA 568 | Db 500 AAAGATTAATGATTTGGATTCATTAAGAAAGGGTAACACGGTTCAAA 559 |
| Qy 569 ATTTGACACTATGCTCTGCTCCACATCTGGGATATTCTGAGATGGGATCATCTGGATCTGGT 628 | Db 560 ATTTGACCCGTTATGCTGCTCCACATCTGGGATATTCTGAGATGGGATCATCTGGT 619 |
| Qy 629 TTACGTTCCACCTCTCTGTGAGTTAACCAAACCTATAAAC 687 | Db 620 TGAGCTCTAACGCTGCCAATTTCGAGTTAACGTTAACCTATAAGT 679 |
| Qy 688 ATCAGCTCTAACAGCTGCCAATTTCGAGCTCTGTGAGTTAACGTTAACCTATAAGT 758 | Db 680 ATCAGCTCTAACAGCTGCCAATTTCGAGCTCTGTGAGTTAACCTATAAGT 751 |
| Qy 747 AGCAGCTTCTAACGCTGCCAATTTCGAGCTCTGTGAGTTAACCTATAAGT 758 | Db 740 TGGGGCTTTT 751 |
| RESULT 7 | |
| LOCUS BQ924401 975 bp mRNA linear EST 20-AUG-2002 | DEFINITION NIH_MGC_18 Homo sapiens cDNA clone IMAGE:6459353 |
| DEFINITION 5', tRNA sequence. | |
| ACCESSION BQ924401 | VERSION 1.1 |
| KEYWORD EST. | SOURCE Homo sapiens (human) |
| ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; | |

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1. (bases 1 to 975)
 NIH_MGC http://mgc.ncbi.nih.gov/
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 COMMENT
 Contact: Robert Strausberg, Ph.D.
 Email: cgbps@small1.nih.gov
 Tissue Procurement: DCDP/DTP/Gazdar
 cDNA Library Preparation: Rubin Laboratory
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LILNL)
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LILNL at:
<http://image.llnl.gov>
 Plate: LILCM264; row: e column: 18
 High quality sequence stop: 661.
 Location/Qualifiers

FEATURES source
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:6459353"
 /tissue_type="large cell carcinoma"
 /lab_host="DH10B (phage-resistant)"
 /clone_id="NIH_MGC_18"
 ECORI: cDNA made by oligo-dT priming. Directionally cloned
 into EcoRI/XbaI sites using the following 5' dddapcr:
 GGACCCAG(G). Library constructed by Ling Hong in the
 laboratory of Gerald M. Rubin (University of California,
 Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and
 SuperScript II RT (Life Technologies). Note: this is a
 NIH_MGC Library."

ORIGIN

| Query | Match | Score | Length | Best Local Similarity | Pred. No. | Mismatches | Indels | Gaps | JOURNAL |
|-------|-------|--|--------|--|-----------|---|--------|---|---------|
| Qy | 59 | 537.2 | 13 | 92.0% | 1..3e-144 | 0 | 43 | 6 | COMMENT |
| Qy | 17 | TRCAACCCGAGGGACCATGTCAGCTGGATAACGGGAGACAACGGTCTCA 76 | 286 | TTCCACCCCGAGGGACCATGTCAGCTGGATAACGGGAGACAACGGTCTCA 345 | 77 | TTCATGAAAGGAAATTCTCTTCCCTATTGTGATGTTGAGTCAATGAGATA 136 | 346 | TTCATGAAAGGAAATTGACTCTCCCTGGACCACTTACTCAATGAGATA 405 | Qy |
| Qy | 286 | TTCCACCCCGAGGGACCATGTCAGCTGGATAACGGGAGACAACGGTCTCA 286 | 77 | TTCATGAAAGGAAATTCTCTTCCCTATTGTGATGTTGAGTCAATGAGATA 136 | 137 | AAACCAAAGAAGTGAATATGACTCTCCCTGGACCACTTACTCAATGAGATA 196 | 406 | AAACCAAAGAAGTGAATATGACTCTCCCTGGACCACTTACTCAATGAGATA 465 | Qy |
| Qy | 197 | GCTGAGCTAAAATCATGCAAAGGCCTTCCTCAATGTTGACTTTGTGACACAG 256 | 466 | GCTGAGCTAAAATCATGCAAAGGCCTTCCTCAATGTTGACTTTGTGACACAG 525 | 257 | AGACAAATCTGTTCTGCTGTTCTGCTCACTGAACTGTTACAGGTTACGGCAGGTC 316 | 526 | AGACAAATCTGTTCTGCTGTTCTGCTCACTGAACTGTTACAGGTTACGGCAGGTC 585 | Qy |
| Qy | 586 | ACCTTCACCTGGGGTCCCTGTGAGCTGCTGAGCTGCTGAGCTGCTGAGCTG 586 | 317 | ACCTTCACCTGGGGTCCCTGTGAGCTGCTGAGCTGCTGAGCTGCTGAGCTG 376 | 586 | ACCTTCACCTGGGGTCCCTGTGAGCTGCTGAGCTGCTGAGCTGCTGAGCTG 645 | 377 | ATGCGAGCTGAGCTTCATGTTGTTCTGCTGAGCTGCTGAGCTGCTGAGCTG 436 | Qy |
| Qy | 646 | ATGGCGAGCTTCATGTTGTTCTGCTGAGCTGCTGAGCTGCTGAGCTG 646 | 437 | CAGCTCAGAACCGAGATGGACTGCTGCTGAGCTGCTGAGCTGCTGAGCTG 495 | 706 | CAGCTCAGAACCGAGATGGACTGCTGCTGAGCTGCTGAGCTGCTGAGCTG 765 | Qy | | |
| Qy | 706 | CAGCTCAGAACCGAGATGGACTGCTGCTGAGCTGCTGAGCTGCTGAGCTG 706 | 496 | AATTCCTCACTGGGGTCCCTGTGAGCTGCTGAGCTGCTGAGCTGCTGAGCTG 554 | Qy | | | | |

| | | | | |
|---|---|---|------------------------------------|---|
| Db | 766 | AATTCCCACTGCCAAAGATACTGACCCCTGGATTCCATTAAAGAAAAGGGTAACA | 825 | Nomura,K., Numazaki,R., Ohno,M., Ohsatto,N., Saito,R., Sakazume,N., Sano,H., Sasaki,D., Sato,K., Shibata,K., Shiraiji,T., Tagami,M., Takeda,Y., Waki,K., Watanuki,A., Muramatsu,M. and Hayashizaki,Y. |
| Qy | 555 | AACTCATTACAAATTGGTACCTATTGTTGCTTG-CPTCACCACTCTG-GTTGAGCTGCGG---ACTAC | 609 | Direct Submission |
| Db | 826 | AACTCGATTCCCAATTGGTACCTATTGCTGCTTCCACCACTCTGGGACTACTGG | 885 | Comparational Analysis of Full-Length Mouse cDNAs Compared with Human Genome Sequences Mamm. Genome. 12, 673-677 (2001) |
| Qy | 610 | TGGACATATCCCTGGTCTCTACAGTCCACC--TCTCTTGAGAGTGTACATGGAT- | 666 | Normalization and subtraction of cap-trap-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000) |
| Db | 886 | AACCTATCCCGGTTCTCTACAGGGCCCTCTTGAGAANGGGTCCATGAAATT | 945 | RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384x multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000) |
| Qy | 667 | -GTTTAAGCAACCTATAAACATCAGCT | 694 | computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001) |
| Db | 946 | GGTTTAAGCCATCTTAACTCTGCT | 975 | cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in RIKEN. Genomic Sciences Center and Genome Science Laboratory in Riken contributed to Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. |
| RESULT | 8 | | | Tissues were provided by David A. Hume (Dept. of Biochemistry and Microbiology/Parasitology Institute for Molecular Bioscience University of Queensland Brisbane Q 4072 Australia) whose assistance we gratefully acknowledge. |
| BY744350 | BY744350 | BY744350 RIKEN full-length enriched, bone marrow macrophage Mus musculus cDNA clone IBS0031F03 5' | 659 bp mRNA linear EST 17-DEC-2002 | Please visit our web site (http://genome.gsc.riken.go.jp) for further details. |
| DEFINITION | BY744350 | BY744350.1 | GI:27170890 | Location/Qualifiers |
| ACCESSION | EST | | | 1. . 659 |
| VERSION | | | | /organism="Mus musculus" |
| KEYWORDS | | | | /mol_type="mRNA" |
| SOURCE | MUS | Mus musculus | (house mouse) | /strain="C57BL/6J" |
| ORGANISM | Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Butheria; Rodentia; Sciurognathi; Muridae; Mus. | | | /db_xref="taxon:10990" |
| REFERENCE | 1 | (bases 1 to 659) | | /clone_id="1830031F03" |
| AUTHORS | Okazaki,Y., Furuno,M., Kasukawa,T., Adachi,J., Bono,H., Kondo,S., Nikaido,I., Osato,N., Saito,R., Suzuki,H., Yamamoto,T., Kyiosawa,H., Yagi,K., Tomaru,Y., Hasegawa,Y., Nogami,A., Schonbach,C., Gojobori,T., Baldarelli,R., Hill,D.P., Bult,C., Hume,D.A., Quackenbush,J., Schrimpf,L.M., Kanapin,A., Matsuda,H., Batyalov,S., Beisel,K.W., Blake,J.A., Bradt,D., Brusic,V., Chothia,C., Corbani,L.E., Cousins,S., Dalla,E., Dragani,T.A., Fletcher,C.F., Forrest,A., Fraer,K.S., Gaasterland,T., Gariboldi,M., Gissi,C., Godzik,A., Gough,J., Grimmond,S., Gustincich,S., Hirokawa,N., Jackson,I.J., Jarvis,E.D., Kanai,A., Kurochkin,I.V., Kawasawa,Y., Kedzierski,R.M., King,B.L., Konagaya,A., Kurochkin,I.V., Lee,Y., Lenhard,B., Lyons,P.A., Maglott,D.R., Maltais,L., Marchionni,L., McKenzie,L., Miki,H., Nagashima,T., Okido,T., Pavon,W.J., Pertea,G., Pesole,G., Petrovsky,N., Pillai,R., Pontius,J.U., Qi,D., Ramachandran,S., Ravasi,T., Reed,J.C., Reed,D.J., Reid,J., Ring,B.Z., Ringwald,M., Scheide,C., Semple,C.A., Setou,M., Shimada,K., Suelama,R., Takenaka,Y., Taylor,M.S., Teasdale,R.D., Tomita,M., Verardo,R., Wagner,L., Wahlestedt,C., Wang,Y., Watanabe,Y., Wellis,C., Wilming,L.G., Wynshaw-Boris,A., Yanagisawa,M., Yang,I., Yang,L., Yuan,Z., Zavolan,M., Zhu,Y., Zimmer,A./, Carninci,P., Hayatsu,N., Hirozane-Kishikawa,T., Konno,H., Nakamura,M., Sakazawa,N., Sato,K., Shiraki,T., Waki,K., Kawai,J., Aizawa,K., Arakawa,T., Fukuda,S., Hara,A., Hashizume,W., Imotani,K., Ishii,Y., Itoh,M., Kaga,Y., Miyazaki,A., Sakai,K., Sasaki,D., Shibata,K., Shinagawa,A., Yasunishi,A., Yoshino,M., Waterston,R., Lander,E.S., Rogers,J., Birney,E. and Hayashizaki,Y. | | /tissue="bone marrow" | |
| COMMENT | | | | /cell_type="macrophage" |
| JOURNAL | | | | /clone_id="RIKEN full-length enriched, bone marrow |
| MEDLINE | | | | /clone_id="macrophage" |
| PUBMED | 12466851 | | | /clone_id="RIKEN full-length cDNA" |
| TITLE | | | | |
| COMMENT | | | | |
| Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center(GSC), Yokohama Institute | | | | |
| The Institute of Physical and Chemical Research (RIKEN) | | | | |
| 1-7-22, Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan | | | | |
| Tel: 81-45-503-9222 | | | | |
| Fax: 81-45-503-9216 | | | | |
| Email: genome-res@gsc.riken.go.jp/ | | | | |
| URL: http://Genome.genomics.riken.go.jp/ | | | | |
| Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Carninci,P., Fukuda,S., Hashizume,W., Hayashida,K., Hirozane,T., Hori,F., Imotani,K., Ishii,Y., Itoh,M., Kaga,Y., Kojima,Y., Kondo,S., Koyano,H., Koya,S., Miyazaki,A., Murakura,M., Nakamura,M., | | | | |
| JOURNAL | | | | |
| MEDLINE | | | | |
| COMMENT | | | | |
| Contact: Yoshihide Hayashizaki | | | | |
| Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center(GSC), Yokohama Institute | | | | |
| The Institute of Physical and Chemical Research (RIKEN) | | | | |
| 1-7-22, Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan | | | | |
| Tel: 81-45-503-9222 | | | | |
| Fax: 81-45-503-9216 | | | | |
| Email: genome-res@gsc.riken.go.jp/ | | | | |
| URL: http://Genome.genomics.riken.go.jp/ | | | | |
| Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Carninci,P., Fukuda,S., Hashizume,W., Hayashida,K., Hirozane,T., Hori,F., Imotani,K., Ishii,Y., Itoh,M., Kaga,Y., Kojima,Y., Kondo,S., Koyano,H., Koya,S., Miyazaki,A., Murakura,M., Nakamura,M., | | | | |
| JOURNAL | | | | |
| MEDLINE | | | | |
| COMMENT | | | | |
| Yoshihide Hayashizaki | | | | |
| Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center(GSC), Yokohama Institute | | | | |
| The Institute of Physical and Chemical Research (RIKEN) | | | | |
| 1-7-22, Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan | | | | |
| Tel: 81-45-503-9222 | | | | |
| Fax: 81-45-503-9216 | | | | |
| Email: genome-res@gsc.riken.go.jp/ | | | | |
| URL: http://Genome.genomics.riken.go.jp/ | | | | |
| Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Carninci,P., Fukuda,S., Hashizume,W., Hayashida,K., Hirozane,T., Hori,F., Imotani,K., Ishii,Y., Itoh,M., Kaga,Y., Kojima,Y., Kondo,S., Koyano,H., Koya,S., Miyazaki,A., Murakura,M., Nakamura,M., | | | | |
| JOURNAL | | | | |
| MEDLINE | | | | |
| COMMENT | | | | |
| Yoshihide Hayashizaki | | | | |
| Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center(GSC), Yokohama Institute | | | | |
| The Institute of Physical and Chemical Research (RIKEN) | | | | |
| 1-7-22, Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan | | | | |
| Tel: 81-45-503-9222 | | | | |
| Fax: 81-45-503-9216 | | | | |
| Email: genome-res@gsc.riken.go.jp/ | | | | |
| URL: http://Genome.genomics.riken.go.jp/ | | | | |
| Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Carninci,P., Fukuda,S., Hashizume,W., Hayashida,K., Hirozane,T., Hori,F., Imotani,K., Ishii,Y., Itoh,M., Kaga,Y., Kojima,Y., Kondo,S., Koyano,H., Koya,S., Miyazaki,A., Murakura,M., Nakamura,M., | | | | |
| JOURNAL | | | | |
| MEDLINE | | | | |
| COMMENT | | | | |
| Yoshihide Hayashizaki | | | | |
| Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center(GSC), Yokohama Institute | | | | |
| The Institute of Physical and Chemical Research (RIKEN) | | | | |
| 1-7-22, Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan | | | | |
| Tel: 81-45-503-9222 | | | | |
| Fax: 81-45-503-9216 | | | | |
| Email: genome-res@gsc.riken.go.jp/ | | | | |
| URL: http://Genome.genomics.riken.go.jp/ | | | | |
| Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Carninci,P., Fukuda,S., Hashizume,W., Hayashida,K., Hirozane,T., Hori,F., Imotani,K., Ishii,Y., Itoh,M., Kaga,Y., Kojima,Y., Kondo,S., Koyano,H., Koya,S., Miyazaki,A., Murakura,M., Nakamura,M., | | | | |
| JOURNAL | | | | |
| MEDLINE | | | | |
| COMMENT | | | | |
| Yoshihide Hayashizaki | | | | |
| Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center(GSC), Yokohama Institute | | | | |
| The Institute of Physical and Chemical Research (RIKEN) | | | | |
| 1-7-22, Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan | | | | |
| Tel: 81-45-503-9222 | | | | |
| Fax: 81-45-503-9216 | | | | |
| Email: genome-res@gsc.riken.go.jp/ | | | | |
| URL: http://Genome.genomics.riken.go.jp/ | | | | |
| Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Carninci,P., Fukuda,S., Hashizume,W., Hayashida,K., Hirozane,T., Hori,F., Imotani,K., Ishii,Y., Itoh,M., Kaga,Y., Kojima,Y., Kondo,S., Koyano,H., Koya,S., Miyazaki,A., Murakura,M., Nakamura,M., | | | | |
| JOURNAL | | | | |
| MEDLINE | | | | |
| COMMENT | | | | |
| Yoshihide Hayashizaki | | | | |
| Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center(GSC), Yokohama Institute | | | | |
| The Institute of Physical and Chemical Research (RIKEN) | | | | |
| 1-7-22, Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan | | | | |
| Tel: 81-45-503-9222 | | | | |
| Fax: 81-45-503-9216 | | | | |
| Email: genome-res@gsc.riken.go.jp/ | | | | |
| URL: http://Genome.genomics.riken.go.jp/ | | | | |
| Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Carninci,P., Fukuda,S., Hashizume,W., Hayashida,K., Hirozane,T., Hori,F., Imotani,K., Ishii,Y., Itoh,M., Kaga,Y., Kojima,Y., Kondo,S., Koyano,H., Koya,S., Miyazaki,A., Murakura,M., Nakamura,M., | | | | |
| JOURNAL | | | | |
| MEDLINE | | | | |
| COMMENT | | | | |
| Yoshihide Hayashizaki | | | | |
| Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center(GSC), Yokohama Institute | | | | |
| The Institute of Physical and Chemical Research (RIKEN) | | | | |
| 1-7-22, Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan | | | | |
| Tel: 81-45-503-9222 | | | | |
| Fax: 81-45-503-9216 | | | | |
| Email: genome-res@gsc.riken.go.jp/ | | | | |
| URL: http://Genome.genomics.riken.go.jp/ | | | | |
| Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Carninci,P., Fukuda,S., Hashizume,W., Hayashida,K., Hirozane,T., Hori,F., Imotani,K., Ishii,Y., Itoh,M., Kaga,Y., Kojima,Y., Kondo,S., Koyano,H., Koya,S., Miyazaki,A., Murakura,M., Nakamura,M., | | | | |
| JOURNAL | | | | |
| MEDLINE | | | | |
| COMMENT | | | | |
| Yoshihide Hayashizaki | | | | |
| Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center(GSC), Yokohama Institute | | | | |
| The Institute of Physical and Chemical Research (RIKEN) | | | | |
| 1-7-22, Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan | | | | |
| Tel: 81-45-503-9222 | | | | |
| Fax: 81-45-503-9216 | | | | |
| Email: genome-res@gsc.riken.go.jp/ | | | | |
| URL: http://Genome.genomics.riken.go.jp/ | | | | |
| Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Carninci,P., Fukuda,S., Hashizume,W., Hayashida,K., Hirozane,T., Hori,F., Imotani,K., Ishii,Y., Itoh,M., Kaga,Y., Kojima,Y., Kondo,S., Koyano,H., Koya,S., Miyazaki,A., Murakura,M., Nakamura,M., | | | | |
| JOURNAL | | | | |
| MEDLINE | | | | |
| COMMENT | | | | |
| Yoshihide Hayashizaki | | | | |
| Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center(GSC), Yokohama Institute | | | | |
| The Institute of Physical and Chemical Research (RIKEN) | | | | |
| 1-7-22, Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan | | | | |
| Tel: 81-45-503-9222 | | | | |
| Fax: 81-45-503-9216 | | | | |
| Email: genome-res@gsc.riken.go.jp/ | | | | |
| URL: http://Genome.genomics.riken.go.jp/ | | | | |
| Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Carninci,P., Fukuda,S., Hashizume,W., Hayashida,K., Hirozane,T., Hori,F., Imotani,K., Ishii,Y., Itoh,M., Kaga,Y., Kojima,Y., Kondo,S., Koyano,H., Koya,S., Miyazaki,A., Murakura,M., Nakamura,M., | | | | |
| JOURNAL | | | | |
| MEDLINE | | | | |
| COMMENT | | | | |
| Yoshihide Hayashizaki | | | | |
| Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center(GSC), Yokohama Institute | | | | |
| The Institute of Physical and Chemical Research (RIKEN) | | | | |
| 1-7-22, Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan | | | | |
| Tel: 81-45-503-9222 | | | | |
| Fax: 81-45-503-9216 | | | | |
| Email: genome-res@gsc.riken.go.jp/ | | | | |
| URL: http://Genome.genomics.riken.go.jp/ | | | | |
| Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Carninci,P., Fukuda,S., Hashizume,W., Hayashida,K., Hirozane,T., Hori,F., Imotani,K., Ishii,Y., Itoh,M., Kaga,Y., Kojima,Y., Kondo,S., Koyano,H., Koya,S., Miyazaki,A., Murakura,M., Nakamura,M., | | | | |
| JOURNAL | | | | |
| MEDLINE | | | | |
| COMMENT | | | | |
| Yoshihide Hayashizaki | | | | |
| Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center(GSC), Yokohama Institute | | | | |
| The Institute of Physical and Chemical Research (RIKEN) | | | | |
| 1-7-22, Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan | | | | |
| Tel: 81-45-503-9222 | | | | |
| Fax: 81-45-503-9216 | | | | |
| Email: genome-res@gsc.riken.go.jp/ | | | | |
| URL: http://Genome.genomics.riken.go.jp/ | | | | |
| Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Carninci,P., Fukuda,S., Hashizume,W., Hayashida,K., Hirozane,T., Hori,F., Imotani,K., Ishii,Y., Itoh,M., Kaga,Y., Kojima,Y., Kondo,S., Koyano,H., Koya,S., Miyazaki,A., Murakura,M., Nakamura,M., | | | | |
| JOURNAL | | | | |
| MEDLINE | | | | |
| COMMENT | | | | |
| Yoshihide Hayashizaki | | | | |
| Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center(GSC), Yokohama Institute | | | | |
| The Institute of Physical and Chemical Research (RIKEN) | | | | |
| 1-7-22, Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan | | | | |
| Tel: 81-45-503-9222 | | | | |
| Fax: 81-45-503-9216 | | | | |
| Email: genome-res@gsc.riken.go.jp/ | | | | |
| URL: http://Genome.genomics.riken.go.jp/ | | | | |
| Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Carninci,P., Fukuda,S., Hashizume,W., Hayashida,K., Hirozane,T., Hori,F., Imotani,K., Ishii,Y., Itoh,M., Kaga,Y., Kojima,Y., Kondo,S., Koyano,H., Koya,S., Miyazaki,A., Murakura,M., Nakamura,M., | | | | |
| JOURNAL | | | | |
| MEDLINE | | | | |
| COMMENT | | | | |
| Yoshihide Hayashizaki | | | | |
| Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center(GSC), Yokohama Institute | | | | |
| The Institute of Physical and Chemical Research (RIKEN) | | | | |
| 1-7-22, Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan | | | | |
| Tel: 81-45-503-9222 | | | | |
| Fax: 81-45-503-9216 | | | | |
| Email: genome-res@gsc.riken.go.jp/ | | | | |
| URL: http://Genome.genomics.riken.go.jp/ | | | | |
| Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Carninci,P., Fukuda,S., Hashizume,W., Hayashida,K., Hirozane,T., Hori,F., Imotani,K., Ishii,Y., Itoh,M., Kaga,Y., Kojima,Y., Kondo,S., Koyano,H., Koya,S., Miyazaki,A., Murakura,M., Nakamura,M., | | | | |
| JOURNAL | | | | |
| MEDLINE | | | | |
| COMMENT | | | | |
| Yoshihide Hayashizaki | | | | |
| Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center(GSC), Yokohama Institute | | | | |
| The Institute of Physical and Chemical Research (RIKEN) | | | | |
| 1-7-22, Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan | | | | |
| Tel: 81-45-503-9222 | | | | |
| Fax: 81-45-503-9216 | | | | |
| Email: genome-res@gsc.riken.go.jp/ | | | | |
| URL: http://Genome.genomics.riken.go.jp/ | | | | |
| Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Carninci,P., Fukuda,S., Hashizume,W., Hayashida,K., Hirozane,T., Hori,F., Imotani,K., Ishii,Y., Itoh,M., Kaga,Y., Kojima,Y., Kondo,S., Koyano,H., Koya,S., Miyazaki,A., Murakura,M., Nakamura,M., | | | | |
| JOURNAL | | | | |
| MEDLINE | | | | |
| COMMENT | | | | |
| Yoshihide Hayashizaki | | | | |
| Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center(GSC), Yokohama Institute | | | | |
| The Institute of Physical and Chemical Research (RIKEN) | | | | |
| 1-7-22, Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan | | | | |
| Tel: 81-45-503-9222 | | | | |
| Fax: 81-45-503-9216 | | | | |
| Email: genome-res@gsc.riken.go.jp/ | | | | |
| URL: http://Genome.genomics.riken.go.jp/ | | | | |
| Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Carninci,P., Fukuda,S., Hashizume,W., Hayashida,K., Hirozane,T., Hori,F., Imotani,K., Ishii,Y., Itoh,M., Kaga,Y., Kojima,Y., Kondo,S., Koyano,H., Koya,S., Miyazaki,A., Murakura,M., Nakamura,M., | | | | |
| JOURNAL | | | | |
| MEDLINE | | | | |
| COMMENT | | | | |
| Yoshihide Hayashizaki | | | | |

| | | | | |
|------------|--|--|-----------------------|---|
| Db | 429 | TTGTAGAGCCAGTCATGAGTCGGATGGCTGGTCTGGAGATTCTACAGATTC | 488 | Nomura, K., Numazaki, R., Ohno, M., Ohsato, N., Saito, R., Sakazume, N., Sano H., Sasaki D., Sato, K., Shibata, K., Shiraki, T., Tagami, M., Takeda, Y., Waki, K., Watanuki, A., Muramatsu, M. and Hayashizaki, Y. |
| Qy | 488 | GTTGAACTTAATCCAACTGAAAGATTACTGAACTTGGATTCAATAAGAAGG | 547 | Direct Submission |
| Db | 489 | GGGAACACATTCCTAACGTCAGTCAGAACTGATATTGGATTCATANGGAAAG | 548 | Computational Analysis of Full-Length Mouse cDNAs Compared with Human Genome Sequences Mamm. Genome 12, 673-677 (2001) |
| Qy | 548 | GTAACAAACTCTGATTCAAAATTGACCTTGGCTCACCATTGGACT | 607 | Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000) |
| Db | 549 | GTAACAAACTCTGATTCAAAATTGACCTTGGCTCACCATTGGACT | 608 | RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384x multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000) |
| Qy | 608 | ACTGGACATATTCCTGGTTCTACAGTTCCACCTCTTGAGGTGTAC | 659 | Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001) |
| Db | 609 | ACTGGACATATTCCTGGTTCTACAGTTCCACCTCTTGAGGTGTAC | 660 | cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genomic Science Laboratory in Riken contributed to Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. |
| RESULT 10 | | | | Tissues were provided by David A. Hume (Dept. of Biochemistry and Microbiology/Parasitology Institute for Molecular Bioscience University of Queensland Brisbane, Q. 4072 Australia) whose assistance we gratefully acknowledge. |
| BY744810 | BY744810 | 666 bp mRNA linear EST 17-DEC-2002 | | Please visit our web site (http://genome.gsc.riken.go.jp/) for further details. |
| DEFINITION | BY744810 | RIKEN full-length enriched, bone marrow macrophage Mus musculus cDNA clone I330048H05 5' | | Location/Qualifiers |
| ACCESSION | BY744810 | | | 1. .666 |
| VERSION | BY744810.1 | | | /organism="Mus musculus" |
| KEYWORDS | EST | | | /mol_type="mRNA" |
| SOURCE | Mus musculus (house mouse) | | | /strain="C57Bl/6J" |
| ORGANISM | Mus musculus | | | /db_xref="taxon:10990" |
| REFERENCE | 1 (bases 1 to 666) | | | /clone="1830048H05" |
| AUTHORS | Okaizaki, Y., Furuno, M., Kasukawa, T., Adachi, J., Bono, H., Kondo, S., Osato, N., Saito, R., Suzuki, H., Murinae; Murinae; Mus. Maltais, L., Osato, N., Tomaru, Y., Hasegawa, Y., Nogami, A., Kyiosawa, H., Yagi, K., Yamamoto, T., Schonbach, C., Gojobori, T., Baldarelli, R., Hill, D. P., Bult, C., Hume, D. A., Quackenbush, J., Schriml, L. M., Kanapin, A., Matsuda, H., Batalov, S., Beisel, K. W., Blake, J. A., Bradt, D., Brusic, V., Chothia, C., Corbani, L. E., Cousins, S., Dalla, E., Drazenic, T. A., Fletcher, C. F., Forrest, A., Friend, K. S., Gaasterland, T., Gariboldi, M., Gissi, C., Godzik, A., Gough, J., Grimmond, S., Gustincich, S., Hirokawa, N., Jackson, I. J., Jarvis, E. D., Kanai, A., Kurochkin, I. V., Kawabe, Y., Kedzierski, R. M., King, B. L., Konagaya, A., Kurochkin, I. V., Lee, Y., Lenhard, B., Lyons, P. A., Maglott, D. R., Maltais, L., Marchionni, L., McKenzie, L., Miki, H., Nagashima, T., Okido, T., Pavani, W. J., Pertea, G., Pesole, G., Petrovsky, N., Pillai, R., Pontius, J. U., Qi, D., Ramachandran, S., Ravasi, T., Reed, J. C., Reed, D. J., Reid, J., Ring, B. Z., Ringwald, M., Ruvolo, R., Schneider, C. C., Semple, C. A., Setou, M., Shimada, K., Sultana, R., Takenaka, Y., Taylor, M. S., Teasdale, R. D., Tomita, M., Well, C., Wilmung, L. G., Wynshaw-Boris, A., Yang, Y., Yang, I., Yang, L., Yuan, Z., Zavolan, M., Zhu, Y., Zimmer, M., Carninci, P., Hayatsu, N., Hirozane-Kishikawa, T., Kishiraki, T., Konno, H., Nakamura, M., Nakamura, N., Sato, K., Shiraishi, T., Waki, K., Kawaji, J., Kizawa, K., Arakawa, T., Fukuda, S., Hara, A., Hashizume, W., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Miyazaki, A., Sakai, K., Sasaki, D., Shibata, K., Shinagawa, A., Yasunishi, A., Yoshino, M., Waterston, R., Lander, E. S., Rogers, J., Birney, E. and Hayashizaki, Y. | | /tissue="bone marrow" | |
| TITLE | | | | /cell_type="macrophage" |
| JOURNAL | Nature | 420, 563-573 (2002) | | /clone_id="RIKEN full-length cDNAs of 60,770 full-length cDNAs |
| MEDLINE | 12466851 | | | 188 ATGACCCAAGCTCAGCTAAATCATGCAAAGCGCCATTCTCTCAATGTGTTG 247 |
| PUBMED | | | | 187 ATGATCTGCCATCAGTAAATCATGCAAATGTGTTGATTCTCTCAAGCTTGTG 246 |
| COMMENT | Contact: Yoshihide Hayashizaki | | | 186 ATGACACAGAGAACAAATCAGTTCGTTGTTGTTGTTGTTGTTGTTG 245 |
| | Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute | | | 185 ATGACACAGAGAACAAATCAGTTCGTTGTTGTTGTTGTTGTTGTTG 244 |
| | The Institute of Physical and Chemical Research (RIKEN) | | | 184 ATGACACAGAGAACAAATCAGTTCGTTGTTGTTGTTGTTGTTGTTG 243 |
| | Te: 81-45-503-9222 | | | 183 ATGACACAGAGAACAAATCAGTTCGTTGTTGTTGTTGTTGTTGTTG 242 |
| | Fax: 81-45-503-9216 | | | 182 ATGACACAGAGAACAAATCAGTTCGTTGTTGTTGTTGTTGTTGTTG 241 |
| | Email: genome-res@gsc.riken.go.jp/ | | | 181 ATGACACAGAGAACAAATCAGTTCGTTGTTGTTGTTGTTGTTGTTG 240 |
| | URL: http://genome-gsc.riken.go.jp/ | | | 180 ATGACACAGAGAACAAATCAGTTCGTTGTTGTTGTTGTTGTTGTTG 239 |
| | Adachi, J., Aizawa, X., Arikawa, T., Arakawa, T., Carninci, P., Fukuda, S., Hashizume, W., Hayashida, K., Hirozane, T., Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Koya, S., Miyazaki, A., Murata, M., Nakamura, M., | | | 179 ATGACACAGAGAACAAATCAGTTCGTTGTTGTTGTTGTTGTTGTTG 238 |

| | | | | |
|------------|--|--|--|--|
| Db | 426 | TGTAGAGCGAGTCATAGTCGATGGCTGGATGGCTGGTCTGCTGAGTCTAGATG | 485 | Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Carninci, P., Fukuda, S., Hashizume, W., Hayashida, K., Hirose, F., Imoto, K., Ishii, Y., Itoh, M., Kegawa, T., Kawai, J., Kojima, Y., Kondo, S., Komio, H., Koya, S., Miyazaki, A., Murata, M., Nakamura, M., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N., Saito, R., Sakazume, N., Sano, H., Sasaki, D., Sato, K., Shiraki, T., Tagami, M., Takeda, Y., Waki, K., Watanabe, M., and Hayashizaki, Y. |
| Qy | 488 | GTTAACCTTAATCCCAACTGGAAAGATTAATGACATTGGATTCGATTTGAGAAGG | 547 | Computational Analysis of Full-Length Mouse cDNAs Compared with Human Genome Sequences Mamm. Genome. 12, 673-677 (2001) |
| Db | 486 | GGAAACAACTATCCTCAACTGCAAAGATCAGTATTTGGATTCCATANGGAAAGG | 545 | Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000) |
| Qy | 548 | GTAACAAACTCGATTCAAAATTGACCTTAATGGCTCTGCTGCCACATCTGGACT | 607 | RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000) |
| Db | 546 | GTAACAAACTCGATTCAAAATTGACCTTAATGGCTCTGCTGCCACATCTGGACT | 605 | Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001) |
| Qy | 608 | ACTGGACATATCCTGGTCTTACAGTCCACCTCTCTTGAGTGTCACATGGATTG | 667 | cDNA library was prepared and sequenced in Mouse Genome Exploration Research Group in RIKEN Genomic Sciences Center and Genome Research Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. |
| Db | 606 | ACTGGACATACCTGGCTCTGAGTCACGTTGACTGTCAGTGTCACTGGATG | 665 | Tissues were provided by David A. Hume (Depts. of Biochemistry and Microbiology/Parasitology Institute for Molecular Bioscience University of Queensland Brisbane Q 4072 Australia) whose assistance we gratefully acknowledge. |
| Qy | 668 | T 668 | Please visit our web site (http://genome.gsc.riken.go.jp/) for further details. | |
| Db | 666 | T 666 | Location/Qualifiers | |
| | | | 1. -656 | /organism="Mus musculus" |
| | | | | /mol_type="mRNA" |
| | | | | /strain="CS7BL/6J" |
| | | | | /db_xref=Taxon:10090 |
| | | | | /clone="I330044H06" |
| | | | | /tissue_type="bone marrow" |
| | | | | /cell_type="macrophage" |
| | | | | /clone_id="RIKEN full-length enriched, bone marrow macrophage" |
| RESULT | 11 | | | |
| BY744678 | | | | |
| LOCUS | BY744678 | 656 bp mRNA linear EST 17-DEC-2002 | | |
| DEFINITION | BY744678 | RIKEN full-length enriched, bone marrow macrophage Mus musculus cDNA clone I830044H06 5', mRNA sequence. | | |
| ACCESSION | BY744678 | | | |
| KEYWORDS | EST. | GI:27171468 | | |
| SOURCE | Mus musculus (house mouse) | | | |
| ORGANISM | Mus musculus | | | |
| REFERENCE | Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Butheria; Rodentia; Sciurognath; Muridae; Murine; Mus. | | | |
| AUTHORS | Okazaki, Y., Furuno, M., Kasukawa, T., Adachi, J., Bono, H., Kondo, S., Niikido, I., Osato, N., Saito, R., Suzuki, H., Yamamoto, I., Kiyosawa, H., Yagi, K., Tomaru, Y., Hasegawa, Y., Nogami, A., Schonbach, C., Gojobori, T., Baldarelli, R., Hill, D. P., Bult, C., Hume, D. A., Quackenbush, J., Schriml, L. M., Kanapin, A., Matsuda, H., Balatov, S., Beisel, K. W., Blake, J. A., Bradt, D., Brusic, V., Chothia, C., Corbani, L. E., Cousins, S., Dalla, E., Dragani, T. A., Fletcher, C. F., Forrest, A., Frazer, K. S., Gaasterland, T., Gustincich, S., Gissi, C., Godzik, A., Gough, J., Grimmond, S., Gariboldi, M., Gissi, C., Jackson, I. J., Jarvis, E. D., Kanai, A., Kawaji, H., Kawasawa, Y., Kedzierski, R. M., King, B. L., Konagaya, A., Kurochkin, I. V., Lee, Y., Lenhard, B., Lyons, P. A., Maglott, D. R., Maltais, L., Marchionni, L., McMenien, L., Miki, H., Nagashima, T., Numata, K., Okido, T., Pavan, W. J., Pertea, G., Pesole, G., Petrovsky, N., Pillai, R., Pontius, J. U., Qi, D., Ramachandran, S., Ravasi, T., Reed, J. C., Reed, D. J., Reid, J., Ring, B. Z., Ringwald, M., Sandelin, A., Schneider, C., Semple, C. A., Setou, M., Shishada, K., Sutaria, R., Takenaka, Y., Terauchi, R., Teasdale, C., Tomita, M., Verardo, R., Wagner, L., Wahlesedt, C., Wang, Y., Watanabe, Y., Wells, C., Wilming, L. G., Wynshaw-Boris, A., Yanagisawa, M., Yang, J., Yang, L., Yuan, Z., Zavolan, M., Zhu, Y., Zimmer, A., Carninci, P., Hayatsu, N., Hirono-Kishikawa, T., Konno, H., Nakamura, M., Sakazume, N., Fukuda, S., Hara, A., Hashizume, W., Imotani, K., Aizawa, K., Itoh, M., Kaga, I., Miyazaki, A., Sakai, K., Shibusawa, K., Shinagawa, A., Yasunishi, A., Yoshino, M., Waterston, R., Lander, E. S., Rogers, J., Birney, E., and Hayashizaki, Y. | | | |
| TITLE | Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs | | | |
| JOURNAL | Nature 420, 563-573 (2002) | | | |
| MEDLINE | 22354683 | | | |
| PUBLMED | 12466651 | | | |
| COMMENT | Contact: Yoshihide Hayashizaki Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute of Physical and Chemical Research (RIKEN) 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan | | | |
| | Fax: 81-45-503-3222 | | | |
| | Email: genome-rs@gscc.riken.go.jp/ | | | |
| | URL: http://genome.gsc.riken.go.jp/ | | | |
| Query | 53.7% | Score 489.6; DB 13; | Length 656; | |
| | Best Local Similarity 85.4%; Pred. No. 7.2e-13; | | | |
| Matches | 554; Conservative 0; Mismatches 94; Indels 1; Gaps 1; | | | |
| Qy | 8 CGGTGAGTTCCAGGGGACCATGTCAGGTGGATAACGGGAGACA | 67 | | |
| Db | 9 CTGCACAGGCTCTCCATCGGAGCATGGCTGAGCTGGGGTACGGCAGCACA | 68 | | |
| Qy | 68 ACCGCTCTTATTCACTGGAGGATTTCCCPATGTGTGATGTCAGCAATCTCAA | 127 | | |
| Db | 69 ACCGCTCCATTCACTGGATGATGTTGTCCTCPATGTGTGATGTCAGTC | 128 | | |
| Qy | 128 TTGAGATTAACCAAGAACAGAAATGAGTCTCCGACCACTTAGTCAAGT | 187 | | |
| Db | 129 TTGAGATTAACCAAGAACAGAAATGAGTCTCCGACCACTTAGTCAAGT | 188 | | |
| Qy | 188 ATGACCCAGCTCACTGGCAACAGCGCCATTCTCAATGTTGACTTGT | 247 | | |
| Db | 189 ATGATCCCTGCCATGCTAACATCAGCAATAGTGGCCATTCACGTTGACTTGT | 248 | | |
| Qy | 248 ATGACACAGGAACTAACTGAGTCGCTGGCTGCTCACTGGAGCTACGGTAC | 307 | | |
| Db | 249 ACCACACGAGGAATACTGATCTGCTGGAGCTCCATGGAACTACGGTGC | 308 | | |
| Qy | 308 GGCGGTCTCACCTTCACTGGGTGATGACCGGCTCCGATAGTATGATG | 367 | | |
| Db | 309 GGAGTTCCATTGCACTGGGGTCAAGAGTATCATGGCTAGGATAGC | 368 | | |
| Qy | 368 GAGTAGCTATGGCAGAGCTCACGTTGACGAAATCCCCAGCT | 427 | | |

| | | | | | | | |
|------------|--|---|--|---------|--|---|-----|
| Db | 369 | GAGTGGAGGTATGCTGGAGAGCTGCATGTGTCACATTGAAATTAGACAAAATAACCCGAGCT | 428 | Qy | 173 | CACTTAGTATCAAAGTATGACCCAAAGGTCAAGCTCAGGTTAAATCATGACAAACAGGGCCATTCCT | 232 |
| Qy | 428 | TGTTGAGGAGCTCATGAAACAGATGGACGGCTCTCTGGAGGTGGTTTACAGATG | 487 | Db | 121 | CTCTTGTTGATCAAAGTACGATGCCAGCTCAGTAATCATGAGTATAAGTCCT | 180 |
| Db | 429 | TGTTGAGGAGCTCATGAAACAGATGGACGGCTCTCTGGAGGTGGTTTACAGATG | 488 | Qy | 233 | TCATGTTGACTTGTGATGACAGAGAACAAATCGTTCTGGCTCTGGCTCCNGAGTATTCAGATG | 292 |
| Qy | 488 | GTGAACCTAATCCCACACTGAAAGATTACTGACACTTGGATTCCATTAGAGTACAGATG | 547 | Db | 181 | TCAACCTTGACTTGTGCGACAGGGACAAATCGTTCTGGCTGGGGCCCTGACTG | 240 |
| Db | 489 | GGAAACAAATCTCAACTGAAAGATCACTGAAAGAAAGG | 547 | Qy | 293 | GAAGCTACAGGTTACCGAGGTACCTTACTGGGGTCCGCTGAGGACAGGGCTCGGTCG | 352 |
| Qy | 548 | GTAAACAAACTGATTCACAAATTGACCTTGTGACTTGTGATTTGATCA-TAAGGAAAGG | 607 | Db | 241 | GAAGCTACAGGTTACCGAGTTACCTTACTGGGGTCACTGAGTACGGCTCTGGACT | 300 |
| Db | 548 | GTAAACAAACTGATTCACAAATTGACCTTGTGACTTGTGATTTGAGTGT | 607 | Qy | 353 | AGCACATAGTAGATGGAGTGGCTATGCTGAGCTTCAATGGAATTTCAGTGAATTTCAG | 412 |
| Qy | 608 | ACTGGACATATCTGGTCTCTTACAGTTGACCTTGTGAGACTGT | 656 | Db | 301 | AACTATGGTGATGGCTGAGGTTACGGTCACTGGTCACTGGAAATTTCAG | 360 |
| Db | 608 | ACTGGACATATCTGGTCTCTTACAGTTGACCTTGTGAGACTGT | 656 | Qy | 413 | ACAAATACCCAGCTTGTGAGGAGCTATGAAACAGGGCTCATGAACTGGACTGGCTGAGTGGAG | 472 |
| Db | 661 | ACAAATACCCAGCTTGTGAGGAGCTATGAACTGGACTGGAGCTATGGAACTGGAA | 420 | Db | 361 | ACAAATACCCAGCTTGTGAGGAGCTATGAACTGGACTGGAACTGGAA | 420 |
| RESULT | 12 | | | Qy | 473 | TGTTTACAGATGGAACTGAAAGATTACTGACACTTGATTGATTGATTGATT | 532 |
| LOCUS | CF768871 | CF768871 | 692 bp mRNA linear EST 17-OCT-2003 | Db | 421 | TATTTTACAGATGGAACTGAAAGATTGATTGATTGATTGATTGATTGATT | 480 |
| DEFINITION | | Bos taurus skin cDNA library Bos taurus cDNA clone | | Qy | 533 | CCATTAAAGAAAAGGTTAAACAAACTGATTCAAAATTGACCTATTGCTCTTC | 592 |
| ACCESSION | CCLO00987 | CCLO00987 | Bos taurus mRNA sequence | Db | 481 | CCATTAAAGAAAAGGTTAAACAAACTGATTGCTCTTC | 540 |
| VERSION | CF768871 | CF768871 | GI:37718090 | Qy | 593 | CACCATCTGGACTACTGGACATA TCCCTGGTTCTTACAGTICACACT-TCCTG | 650 |
| KEYWORDS | BST. | Bos taurus (cow) | | Db | 541 | CTCCATGCCGGATATTGGACATATTCCCTGGCTCTTACGGTCACTCTGGAA | 600 |
| SOURCE | | Bos taurus | | Qy | 651 | GAGTGTAC-ATGGATGTTAAACCAACCTTAAACATCA 691 | |
| ORGANISM | | Bos taurus; Chordata; Craniata; Vertebrata; Buteleostomi; | | Db | 601 | AAGCGTACATTGGATCATTTAAAGCAATCCATAACATTA 642 | |
| REFERENCE | 1 (bases 1 to 692) | | | | | | |
| AUTHORS | Wang, Y.H., McWilliam, S. and Lehner, S. | | | | | | |
| TITLE | Transcription profiling of cattle skin | | | | | | |
| JOURNAL | Unpublished (2003) | | | | | | |
| COMMENT | Contact: Dr Yonghong Wang Functional Genomics Lab CSIRO Livestock Industries Level 5, Queensland and Biosciences Precinct, University of Queensland, 3061 Carmody Road St. Lucia QLD Australia | | | | | | |
| FEATURES | source | | | | | | |
| | | LOCUS BY22663 | BY22663 | BY22663 | 639 bp mRNA full-length enriched, adult male bone Mus musculus | EST 17-DEC-2002 | |
| | | DEFINITION BY22663 RIKEN full-length enriched, adult male bone Mus musculus | BY22663 | BY22663 | cdNA clone 9830112C14 5', mRNA sequence. | | |
| | | ACCESSION BY22663 | BY22663 | BY22663 | EST. | | |
| | | VERSION BY22663.1 | GI:27135780 | | | | |
| | | KEYWORDS | | | | | |
| | | SOURCE | Mus musculus (house mouse) | | | | |
| | | ORGANISM | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Buteleostomi; | | | | |
| | | REFERENCE | Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. | | | | |
| | | AUTHORS | 1 (bases 1 to 639); Furuno, M., Kasukawa, T., Adachi, J., Bono, H., Kondo, S., Nikaido, I., Osato, N., Saito, R., Suzuki, H., Yamakawa, I., Kiyosawa, H., Yagi, K., Tomaru, Y., Hasegawa, Y., Nagami, A., Schonbach, C., Gojobori, T., Baldarelli, R., Hill, D.P., Bult, C., Hume, D.A., Quackenbush, J., Schriml, L.M., Kanapin, A., Matsuda, H., Batyalov, S., Beisel, K.W., Bradt, J.A., Brusic, V., Chothia, C., Corbani, L.E., Cousins, S., Dalla, R., Dragani, T.A., Fletcher, C.F., Forrest, A., Frazer, K.S., Gaasterland, T., Garibaldi, M., Gissi, C., Godzik, A., Gough, J., Grimes, J., Gusiussev, S., Hirokawa, N., Jackson, I.J., Jarvis, E.D., Kanai, A., Kawaji, H., Kawasawa, Y., Kedzierski, R.M., King, B.L., Konagaya, A., Kurochkin, I.V., Lee, Y., Lenhard, B., Lyons, P.A., Maglott, D.R., Maltais, L., Marchionni, L., McKenzie, L., Miki, H., Nasashima, T., Numata, X., Okido, T., Pavani, W.J., Perera, G., Pesce, G., Petrovsky, N., Pilai, R., Portius, J.U., Qi, D., Ramachandran, S., Rayasi, T., Reed, J.C., Reid, D.J., Reid, J.C., Ringwald, M., Sandelin, A., Schneider, C., Semple, C.A., Setou, M., Shimada, K., Sultana, R., Takenaka, Y., Taylor, M.S., Teasdale, R.D., Tomita, M., Verardo, R., Wagner, L., Wahlestedt, C., Wang, Y., Watanabe, Y., Wells, C., Wilming, L.G., Wynshaw-Boris, A., Yanagisawa, M., Yang, J., Yang, L., Yuan, Z., Zavolan, M., Zhu, Y., Zimmer, A., Carninci, P., Hayatsu, N., Hirozane-Kishikawa, T., Konno, H., Nakamura, M., | | | | |
| | | | | | | | |
| | | Best Local Similarity 87.1%;保守性 55.9%;保守性 | Score 478; DB 14; Length 692; | | | | |
| | | Matches 0; Mismatches 80; Indels 3; Gaps 3; | Pred. No. 1.8-127; | | | | |
| | | | | | | | |
| Qy | 52.5* | | | | | | |
| Db | 53 | GATACCGGGAGCACCGCTTATTCAGGATATTCCTTATGTTGATGGTG | 112 | Qy | 53 | GATACCGGGAGCACCGCTTATTCAGGATATTCCTTATGTTGATGGTG | 60 |
| Db | 1 | GETACTCGAGGCCGAGCTCTGTTATGTTGATGGTG | 60 | Db | 61 | ATCAGCAATCTCCAATTGAGATAAAACCAAGAATGTAATTGACTCTCCCTCGAC | 172 |
| Qy | 113 | ATCAGCAATCTCCAATTGAGATAAAACCAAGAATGTAATTGACTCTCCCTCGAC | 120 | db | 61 | ATCAACATCTCGATGAGATAAAAGAATGAGATGATTGATGATTCTCCCTCGAC | 120 |

Sakazume, N., Sato, K., Shiraki, T., Waki, K., Kawai, J., Aizawa, K., Arakawa, T., Fukuda, S., Hara, A., Hashizume, W., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Miyazaki, A., Sakai, K., Sasaki, D., Shiba, K., Shinagawa, A., Yasunishi, A., Yoshino, M., Waterston, R., Lander, E.S., Rogers J., Birney, E. and Hayashizaki, Y.

TITLE Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs

JOURNAL Nature 420, 563-573 (2002)

MEDLINE 124466851

COMMENT Contact: Yoshihide Hayashizaki Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute The Institute of Physical and Chemical Research (RIKEN) 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan Tel: 81-45-503-9222 Fax: 81-45-503-9216

URL: <http://genome-res.gsc.riken.go.jp/>

Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Carninci, P., Fukuda, S., Hashizume, W., Hayashida, K., Hirozane, T., Hori, F., Imotani, K., Ishii, Y., Ichio, M., Kaga, I., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Koya, S., Miyazaki, A., Murata, M., Nakamura, M., Nomura, K., Numazaki, R., Ohno, M., Obsato, N., Saito, R., Sakazume, N., Sano, H., Sasaki, D., Sato, K., Shibata, K., Shiraki, T., Tagami, M., Takeda, Y., Waki, K., Watshiki, A., Muramatsu, M. and Hayashizaki, Y.

Direct Submission

Computational Analysis of Full-Length Mouse cDNAs Compared with Human Genome Sequences Mamm. Genome 12, 673-677 (2001)

Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (10), 1617-1630 (2000)

RIKEN integrated sequence analysis (RISA) system -384-format sequencing pipeline (2000)

Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)

CDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. Please visit our web site (<http://genome.gsc.riken.go.jp/>) for further details.

FEATURES**source**

```

1. .639
  /organism="Mus musculus"
  /mol_type="mRNA"
  /db_xref="Taxon:10090"
  /clone="9830112C14"
  /sex="male"
  /tissue_type="bone"
  /dev_stage="adult"
  /lab_host="DH10B"
  /clone_lib="RIKEN full-length enriched, adult male bone"
  /note="Site 1: Sali; Site 2: BamHI; cDNA library was
  prepared and sequenced in Mouse Genome Encyclopedia
  Project of Genome Exploration Research Group in Riken
  Genomic Sciences Center and Genome Science Laboratory in
  RIKEN. Division of Experimental Animal Research in Riken
  contributed to prepare mouse tissues. 1st strand cDNA was
  primed with a primer [5,
  GAGAGAGAGAGATCCAAGAGCTCTTTRTTTTRVNN 3'], cDNA was
  prepared by using triehalose thermo-activated reverse
  transcriptase and subsequently enriched for full-length by
  cap-trapper. cDNA went through one round of normalization
  to Rot = 10.0 and subtraction to Rot = 185.0. Second
  strand cDNA was prepared with the primer adapter of
  sequence [5', GAGAGAGAGATTCGAATTAAATTAAATCCCCCCCCCCC
  3']. cDNA was cleaved with XbaI and BamHI. Vector: a Lambda
  modified pBluescript KS(+) after bulk excision from Lambda
  plC I."

```

| ORIGIN | |
|--------|---|
| Query | Match 49.5%; Best Local Similarity 85.3%; Matches 501; Conservative 0; Mismatches 86; Indels 0; Gaps 0; |
| Db | 8 CGCGTCAGITCCACCCGGGACCATGTCAGGTGATACTCTCAA 53 CTGCACAGGCTCTCCATGAGGCCATGGGAGTGGCTGAGCTGAGCA 112 |
| Qy | 68 ATGGCTCTATTACTGAAAGGATTTCCTTATGTTGATGAGCAATCTCAA 127 113 ACCTGCCATTCTGAAATGATGTTGATGAGCTCTCAA 172 |
| Db | 128 TTGAGNTAAACCACAGACTGAAATATGACTTCCTCGAACCACTTGTAAGT 187 173 TTGAGATTAACCAARGAAGTGAATAAGACTCTCACTCGAACCTCTCACTATCAAGT 232 |
| Qy | 188 ATGACCAAGCTGCTAAATCATCGAAACAGGGCATCTCAATGTTGACTTG 247 233 ATGATCTGCCCTGAGTAAATCATGCAATTAGTGCCTCAACGGTGAATTTG 292 |
| Db | 248 ATGACACAGAGACAAATCAGTTCTGGTGTGGTCTCACTGGAGGTACAGGTTCAC 307 293 AGCACAGGAGAACATCAGTTCTGGTGGGTCTCACTGGAAACTACAGGTGC 352 |
| Qy | 308 GCGAGGTCACTTCACTGGGGTCCGCTGATGACCAAGGGCTCCGAGCACATAGTAGATG 367 353 GGAGATTCATTGCACTGGGTCACTGGGTACAGATGATCATGGTCAGAGTAAAG 412 |
| Db | 368 GAGTGAGCTATGCTGGAGGCTCCATTTGTCATCGSGAATTCGCCGAGCT 427 413 GACTGAGGTATGTCAGGAGCTGCACTGGTCATGGTCACTGSGAATTCGCCGAGCT 472 |
| Qy | 428 TTCTGAGGAGCTCATGAAACGAGTGGACTGCTGCTCTGGGATGTTTACAGATTTG 487 473 TTGTAGGGAGCTCATGAGTCGGATGGCTCTGNGAGTATNTCTACAGATTTG 532 |
| Db | 488 GTGAACCTAATTCCAACTGCAAAGATTACTGACACTTTGGATTCATTAAAGAAAAGG 547 533 GGAAACAATTCACAGTCACTGGCAAAAGATCACTGATAATTGGATTCATAAGGAAAGG 592 |
| Qy | 548 GAAACAAACTGATTCACAAATTGGCTATTGCTCTGTCTCC 594 593 GTPAACAAACCGGTCACAAATTGGCTATTGCTCTGTCTCC 639 |
| Db | |

location/Qualifiers

| RESULT 14 | |
|------------|---|
| LOCUS | BB237415 |
| DEFINITION | BE237415 MARC 4Bov Bos taurus cDNA 5', mRNA sequence. |
| ACCESSION | BE237415 |
| KEYWORDS | EST. |
| SOURCE | Bos taurus (cow) |
| ORGANISM | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Buteleostomi; Mammalia; Eutheria; Bovinae; Bovidae; Bovidae; Bovine; Bos. |
| REFERENCE | 1 (bases 1 to 561) |
| AUTHORS | Smith, T. P. L., Grosse, W. M., Freking, B. A., Roberts, A. J., Stone, R. T., Casas, E. B., Wray, J. B., White, J., Cho, J., Fahrenkrug, S. C., Bennett, G. L., Heaton, M. P., Laegreid, W. W., Rohrer, G. A., Chitko-McKown, C. G., Periera, G., Holt, I., Karamycheva, S., Liang, F., Quackenbush, J., and Keele, J. W. |
| TITLE | Sequence evaluation of four pooled-tissue normalized bovine cDNA libraries and construction of a gene index for cattle |
| JOURNAL | Genome Res. 11 (4), 626-630 (2001) |
| MEDLINE | 21180013 |
| PUBMED | 1128978 |
| COMMENT | Contact: Smith TPL USDA, ARS, US Meat Animal Research Center PO Box 166, Clay Center, NE 68933-0166, USA |

Tel: 402 762 4366
 Fax: 402 762 4390
 Email: smith@email.marc.usda.gov
 Single pass sequencing. Bases called and alt_trimmed with phred v0.980924.e. Vector identified by cross_match with the -minscore 18 and -minmatch 12 options.
 PCR Primers
 FORWARD: AGGAAACAGGTATGACCAT
 BACKWARD: GTTTTCCCAAGTCAGCAGC
 Plate: 47 row: F column: 16
 Seq primer: ATTAGCTGACACTATAG
 Location/Qualifiers
 1. .561
 /organism="Bos taurus"
 /mol_type="mRNA"
 /db_xref="taxon:9913"
 /tissue_type="Pooled"
 /lab_host="DHB0B"
 /clone_lab="MARC 4B0V"
 /note="Vector: PCMV SPORT6; Site 1: NotI; Site 2: SalI;
 Library made from pooled tissue From day 20 and day 40
 embryos."
 embryo."

| IGIN | Query Match | Score | DB 10; | Length | 561; |
|--------------------|--|--------|---------------------|--------|-------------------|
| Best Local Matches | Similarity | 49.0% | Score 446.6; | DB 10; | Length 561; |
| Local Matches | Conservative | 88.3%; | Pred. No. 2-2e-118; | | |
| Matches | | 0; | Mismatches | 65; | Indels 1; Gaps 1; |
| 47 | GCTGGGATAACCGGAGGACAAAGGTCTTATCGAAGGATTTCCTATGTTGCTG | 106 | | | |
| 1 | GCTGGGCTACCGGAGGGCAGGTCTTATGTTGCTGTTATGAAATGTTTCCTATGTTG | 60 | | | |
| 107 | ATGGTGATCTAGCAATCTCCAAATGAGATTAAACCAAGAAGTGAATATGACTCTTCCC | 166 | | | |
| 61 | ATGGGATCAAACATCTCGATTTGAGATTAAACCAAGAAGTGAATATGACTCTTCCC | 120 | | | |
| 167 | TCCGACCACTTAACTCAAGTATGAGTTGACCCAAAGCTAGCTAAAATCATCGAAAGCGGCC | 226 | | | |
| 121 | TCCGACCTCTTGTATCAAGTACGTTACGTAATGCAAGTCACTGAAATATGTTGCGC | 180 | | | |
| 227 | ATTCTCTCAATGTTGACTTTGATGACAAGAGAACAAATCAGTCTGCTGTTGCTC | 286 | | | |
| 181 | ACTCTCTAACGTTGACTTTGACGACACGGACGACAAATCAGTCTGTTGGGGCCCC | 240 | | | |
| 287 | TOACTGGAAAGCTACAGTTACGGTTACGGTACGGTACCTTCACTGGGTTCCCTGATGCCAAG | 346 | | | |
| 241 | TOACTGAAAGCTACAGTTACGGTTACGGTACGGTACCTTCACTGGGTTCCCTGATGCCAAG | 300 | | | |
| 347 | GCTCCGACCACTATGAGTATGAGTTGAGCTGAGCTATGCTGCAGAGETCTCACTGGTCACTGGA | 406 | | | |
| 301 | GCTCTGACATGTGGTGGATGGGTGACCTGGAGGTACCTGGAGCTCACCTGGTCACTGGTCACTGGA | 360 | | | |
| 407 | ATTCAAGCAAATACCCAGCTGCTTGTGAGGCAACTCATGAACTTGTGAGCTGACTGCT | 466 | | | |
| 361 | ATTCAAGCAAATACCCAGCTGCTTGTGAGGCAACTCATGAACTTGTGAGCTGACTGCT | 420 | | | |
| 467 | TGGGAGTGTITTAACAGATGGTGAACCTAATCCAACTGCAAAGTATTGACACTT | 526 | | | |
| 421 | TGGGATATTTCAGATGTTGAGCTAAACACTGCTTACAAATTTGACCTATTGCTC | 479 | | | |
| 527 | TGGATTCATTAAAGAAAAGGTTAACAAACACTGCTTACAAATTTGACCTATTGCTC | 586 | | | |
| 480 | TGGATTCATTAAAGAAAAGGTTAACAAACACTGCTTACGAATNTGACCCGGATGTC | 539 | | | |
| 587 | TGCTTCCACCATCTGGACTA | 608 | | | |
| 540 | TACTCCCTCCATGCGGGATTA | 561 | | | |

5' , mRNA sequence.
 ACCESSION BU169155
 VERSION BU169155.1 GI:22683139
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Bokaryota; Metazoa; Chordata; Craniata; Vertebrata; Buteleo
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 ('bases 1 to 928)
 AUTHORS NIH-NCI http://mgc.nci.nih.gov/
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 JOURNAL Contact: Robert Strausberg, Ph.D.
 COMMENT Email: cgs@psr@mail.nih.gov
 Tissue Procurement: ATCC
 cDNA Library Preparation: Rubin Laboratory
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution by: MGC clone distribution information can
 found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 Plate: LLM2312 row: 1 column: 18

Qy 437 CAGCTCATGAAACCAAGATGACTGGC--TGTCCTGGAGTGTCTTACAGATTGGTG-A 491
Db 705 CACCTCATGGAAACAATGACTGGCCTGTCCTGGAAATGTCCTTACAGATGGCCAA 764

Qy 492 ACCTAATCCCAACTGCAAAA 512
Db 765 ACCATATCCCAACTGCAAA 785

Search completed: September 14, 2004, 00:31:32
Job time : 2477 secs

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------|--|
| 1 | 284.8 | 311.3 | 605 | 1 | US-08-459-667-17 Sequence 17, Appli |
| 2 | 284.8 | 311.3 | 605 | 4 | US-09-224-110-17 Sequence 17, Appli |
| 3 | 284.8 | 311.3 | 605 | 5 | PCT-US95-07289-17 Sequence 17, Appli |
| 4 | 280 | 30.7 | 780 | 3 | US-08-921-128-3 Sequence 11, Appli |
| 5 | 280 | 30.7 | 864 | 2 | US-08-595-868C-11 Sequence 11, Appli |
| 6 | 280 | 30.7 | 867 | 3 | US-09-139-819A-11 Sequence 11, Appli |
| 7 | 280 | 30.7 | 867 | 4 | US-09-750-913C-11 Sequence 11, Appli |
| 8 | 278.4 | 30.6 | 1759 | 4 | US-09-976-594-643 Sequence 643, App |
| 9 | 278.4 | 30.6 | 1755 | 4 | US-09-566-921-124 Sequence 124, App |
| 10 | 266.8 | 29.3 | 1266 | 4 | US-09-938-270B-2 Sequence 2, Appli |
| 11 | 177.4 | 19.5 | 599 | 3 | US-09-395-982-446 Sequence 446, App |
| 12 | 147.6 | 16.2 | 477 | 4 | US-09-621-976-18059 Sequence 18059, A |
| 13 | 147.4 | 16.2 | 581 | 3 | US-09-395-982-136 Sequence 136, App |
| c 14 | 136.8 | 15.0 | 760 | 3 | US-09-395-982-156 Sequence 156, App |
| c 15 | 111 | 12.2 | 586 | 3 | US-09-385-982-472 Sequence 472, App |
| 22 | 86.6 | 10.2 | 55298 | 4 | US-09-491-3566C-1 Sequence 1, Appli |
| c 17 | 88.2 | 9.7 | 2721 | 3 | US-08-921-195-1 Sequence 1, Appli |
| c 18 | 88 | 9.7 | 45716 | 4 | US-08-965-048-5 Sequence 5, Appli |
| c 19 | 88 | 9.7 | 45989 | 4 | US-08-965-048-6 Sequence 6, Appli |
| c 20 | 87.8 | 9.6 | 35050 | 3 | US-08-814-095-7 Sequence 7, Appli |
| c 21 | 86.8 | 9.5 | 66804 | 4 | US-09-740-041-3 Sequence 3, Appli |
| c 22 | 86.6 | 9.5 | 1624 | 2 | US-08-852-807-10 Sequence 10, Appli |
| c 23 | 85.6 | 9.5 | 13674 | 2 | US-08-852-807-11 Sequence 1, Appli |
| c 24 | 86.6 | 9.5 | 39982 | 4 | US-09-822-924-3 Sequence 3, Appli |
| c 25 | 86.6 | 9.5 | 161652 | 4 | US-09-497-855A-40 Sequence 40, Appli |
| c 26 | 86 | 9.4 | 392000 | 4 | US-10-027-983-11 Sequence 11, Appli |
| c 27 | 85.8 | 9.4 | 18073 | 3 | US-09-078-294-12 Sequence 12, Appli |

GenCore version 5.1.6
 Copyright (c) 1993 - 2004 Compugen Ltd.
 OM nucleic - nucleic search, using sw model
 Run on: September 13, 2004, 22:29:06 ; Search time 96 Seconds
 (without all alignments)
 Title: US-10-069-434-4
 Perfect score: 911
 Sequence: 1 cggaaatccggctcgaggatcc.....ccccaaatgccaggaaaaa 911
 Scoring table: IDENTITY NUC
 Gapop 10.0 , Gapext 1.0
 Searched: 682709 seqs, 277475446 residues
 Total number of hits satisfying chosen parameters: 1365418

ALIGNMENTS

RESULT 1
 US-08-469-667-17
 ; Sequence 17, Application US/08469667
 ; Patent No. 5737448
 ; GENERAL INFORMATION:
 ; APPLICANT: Yu, Guo-Liang
 ; TITLE OF INVENTION: Colon Specific Genes and Proteins
 ; NUMBER OF SEQUENCES: 24
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Carella, Byrne, Bain, Gilfillan, Cecchi,
 ; STREET: 6 Becker Farm Road
 ; CITY: Roseland
 ; STATE: NJ
 ; COUNTRY: USA
 ; ZIP: 07068-1739
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: FLOPPY disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent in Release #1.0 Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/469, 667
 ; FILING DATE: 06-JUN-1995
 ; CLASSIFICATION: 536
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Portera, Gregory D
 ; REGISTRATION NUMBER: 36,134
 ; REFERENCE/DOCKET NUMBER: 325800-435
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 201-994-1700
 ; TELEFAX: 201-994-1744
 ; INFORMATION FOR SEQ ID NO: 17:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 605 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: cDNA
 ; FEATURE:
 ; NAME/KEY: CDS
 ; LOCATION: 1 .. 603
 ; FEATURE:
 ; NAME/KEY: mat peptide
 ; LOCATION: 1 .. 603
 ; US-08-469-667-17

* Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------|--|
| 1 | 284.8 | 311.3 | 605 | 1 | US-08-459-667-17 Sequence 17, Appli |
| 2 | 284.8 | 311.3 | 605 | 4 | US-09-224-110-17 Sequence 17, Appli |
| 3 | 284.8 | 311.3 | 605 | 5 | PCT-US95-07289-17 Sequence 17, Appli |
| 4 | 280 | 30.7 | 780 | 3 | US-08-921-128-3 Sequence 11, Appli |
| 5 | 280 | 30.7 | 864 | 2 | US-08-595-868C-11 Sequence 11, Appli |
| 6 | 280 | 30.7 | 867 | 3 | US-09-139-819A-11 Sequence 11, Appli |
| 7 | 280 | 30.7 | 867 | 4 | US-09-750-913C-11 Sequence 11, Appli |
| 8 | 278.4 | 30.6 | 1759 | 4 | US-09-976-594-643 Sequence 643, App |
| 9 | 278.4 | 30.6 | 1755 | 4 | US-09-566-921-124 Sequence 124, App |
| 10 | 266.8 | 29.3 | 1266 | 4 | US-09-938-270B-2 Sequence 2, Appli |
| 11 | 177.4 | 19.5 | 599 | 3 | US-09-395-982-446 Sequence 446, App |
| 12 | 147.6 | 16.2 | 477 | 4 | US-09-621-976-18059 Sequence 18059, A |
| 13 | 147.4 | 16.2 | 581 | 3 | US-09-395-982-136 Sequence 136, App |
| c 14 | 136.8 | 15.0 | 760 | 3 | US-09-395-982-156 Sequence 156, App |
| c 15 | 111 | 12.2 | 586 | 3 | US-09-385-982-472 Sequence 472, App |
| 22 | 86.6 | 10.2 | 55298 | 4 | US-09-491-3566C-1 Sequence 1, Appli |
| c 17 | 88.2 | 9.7 | 2721 | 3 | US-08-921-195-1 Sequence 1, Appli |
| c 18 | 88 | 9.7 | 45716 | 4 | US-08-965-048-5 Sequence 5, Appli |
| c 19 | 88 | 9.7 | 45989 | 4 | US-08-965-048-6 Sequence 6, Appli |
| c 20 | 87.8 | 9.6 | 35050 | 3 | US-08-814-095-7 Sequence 7, Appli |
| c 21 | 86.8 | 9.5 | 66804 | 4 | US-09-740-041-3 Sequence 3, Appli |
| c 22 | 86.6 | 9.5 | 1624 | 2 | US-08-852-807-10 Sequence 10, Appli |
| c 23 | 85.6 | 9.5 | 13674 | 2 | US-08-852-807-11 Sequence 1, Appli |
| c 24 | 86.6 | 9.5 | 39982 | 4 | US-09-822-924-3 Sequence 3, Appli |
| c 25 | 86.6 | 9.5 | 161652 | 4 | US-09-497-855A-40 Sequence 40, Appli |
| c 26 | 86 | 9.4 | 392000 | 4 | US-10-027-983-11 Sequence 11, Appli |
| c 27 | 85.8 | 9.4 | 18073 | 3 | US-09-078-294-12 Sequence 12, Appli |

Query Match Best Local Similarity 31.3% ; Score 284.8; DB 1; Length 605; Pred. No. 3.8e-84;

RESULT 2
US 09-224-110-17
Sequence 17, Application US/09224110
Patent No. 6337195
GENERAL INFORMATION:
APPLICANT: Yu, Guo-Liang
TITLE OF INVENTION: Colon Specific Genes and Proteins
NUMBER OF SEQUENCES: 24
CORRESPONDENCE ADDRESS:
ADDRESSEE: Carella, Byrne, Bain, Gilfillan, Cecchi,
ADDRESS: Stewart & Olstein
STREET: 6 Becker Farm Road
CITY: Rosecland
STATE: NJ
COUNTRY: USA
ZIP: 07068-1739
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC DOS/MS-DOS
SOFTWARE: Patienten Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/224,110
FILING DATE:
CLASSIFICATION:
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: 08/469,667
FILING DATE: 06-JUN-1995
ATTORNEY/AGENT INFORMATION:
TITLE OF INVENTION: Colon Specific Genes and Proteins

NAME: Ferraro, Gregory D.
REGISTRATION NUMBER: 36,134
REFERENCE/DOCKET NUMBER: 325800-435
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201-994-1700
TELEX: 201-994-1744
INFORMATION FOR SEQ ID NO: 17:
SEQUENCE CHARACTERISTICS:
LENGTH: 605 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cdna
FEATURE:
NAME/KEY: CDS
LOCATION: 1...603
FEATURE:
NAME/KEY: mat_Peptide
LOCATION: 1...603
US-09-224-110-17

Query Match Score 284.8; DB 4; Length 605;
Best Local Similarity 67.2%; Pred. No. 3..8e-184;
Matches 403; Conservative 0; Mismatches 197; Indels 0; Gaps 0;

Qy 128 TTGGATTAACCAAAAGTCAAATATGACTCTTCCTCCGACCACTTAGTCAAGT 187
Db 2 TTGATATAAACACGTAAACCAACGTTCTGCTGTTTACAGATGT 61

Qy 188 ATGACCCAGCTAGCTAAATCATCGTCTGCCTCTCACTGGAACTCACGGTTAC 247
Db 122 AGGACAAAGATAACCGATACGTGTGAAGGTGTCCTCTGACGCTACAGCT 181

Qy 308 GGAGGTTTACCTTCACTGGGTCGGTGTGACCGGTCGAGACATAGATAGT 367
Db 182 TTCTGTTCAATTCACTGGGCACTGGGAGTAGTAAATGAGCTAGTGGATG 241

Qy 368 GAGTGAGCTATGCTGAGAGCTCCATGTGTCACTGGAAATTAGACAATAACCCACT 427
Db 242 GAGTCAAATATTGCCAGGCTTACGTGGCTTACGTGAATTCGAGAATCTGCACCC 301

Qy 428 TTGTGAGGCACTGGAGCTCATGAAACCAGATGAACTGGCTTGGAGTTCTGTTACAGATGT 487
Db 302 TTGCTGAAGCTGCTCAAGGCTTATGGTTGGAGTTATGGTTGATGAAGGTG 361

Qy 488 GTGAGCTATTCCAACACTGCAAAAGTAACTGACACTGGATTCCATTAAAGAAGG 547
Db 362 GTGAGGCCAACCCAAAGCTGCAAGAAGTACTGGCTCCAGAACATTAAACCAAGG 421

Qy 548 GTAAACAAACTCGATTCAAAUTTGACCTATTGCTTCACTCTGGGACT 607
Db 422 GCACAGGAGCCCCATTCAACATTGACCTTCACTCTCCCTCTCCCTGT 481

Qy 608 ACTGGACATATCCCTGGTTCTCTTACAGTTCACCTCTCTTGAGAGTGTCACTGGATG 667
Db 482 TCTGGACATACCCCTGGCTCTGACTCACTCTCTTATGAGGTGAACCTGGATCA 541

Qy 668 TTTTAAAGGAACTTATAACATCGCTTCACAGCTGCCAAATTGGCAGCTCCGT 727
Db 542 TCTGTTAAGGAGGACATCAGTGTCAAGGAGTTGGCACATTCCGGAGCTCTPAT 601

Qy 724 GAGTCAAATATTGCCAGGCTTACGTGGTCACTGGAAATTGAGAAATAACCCAGT 427
Db 242 GAGTCAAATATTGCCAGGCTTACGTGGTCACTGGAAATTGAGAAATAACCCAGT 301

Qy 428 TTGTGAGGCACTGGCTCATGAAACCAGTGAAGTGGCTTTCAGATGGT 487
Db 302 TTGCTGAAGCTGCTCAAGGCTTATGGTTGATGAAGGTG 361

Qy 488 GTGACCTTAATTCCAACCTGCAAAAGATTACTGACACTGGATTCCTTAANGAAAGG 547
Db 362 GTGAGGCCAACCCAAAGCTGCAAGAAGTACTGGCTTCACTGGGACT 421

Qy 548 GTAAACAAACTCGATTCAAAUTTGACCTATTGCTCAGTCCACCTGGGACT 607
Db 422 GCACAGGAGCCCCATTCAACATTGACCTTCACTCTCCCTCTCCCTGT 481

Qy 608 ACTGGACATATCCCTGGTTCTCTTACAGTTCACCTCTCTTGAGAGTGTCACTGGATG 667
Db 482 TCTGGACATACCCCTGGCTCTGACTCACTCTCTTATGAGGTGAACCTGGATCA 541

Qy 668 TTTTAAAGGAACTTATAACATCGCTTCACAGCTGCCAAATTGGCAGCTCCGT 727
Db 542 TCTGTTAAGGAGGACATCAGTGTCAAGGAGTTGGCACATTCCGGAGCTCTPAT 601

RESULT 3
PCT-US95-07289-17
Sequence 17, Application PC/TUS9507289
GENERAL INFORMATION:
APPLICANT: Yu, Guo-Liang
TITLE OF INVENTION: Colon Specific Genes and Proteins
NUMBER OF SEQUENCES: 24
CORRESPONDENCE ADDRESS:
ADDRESSEE: Carella, Byrne, Bain, Gilfillan, Cecchi,
ADDRESS: Stewart & Olstein
STREET: 6 Becker Farm Road
CITY: Rosecland
STATE: NJ
COUNTRY: USA
ZIP: 07068-1739
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC DOS/MS-DOS
SOFTWARE: Patienten Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/224,110
FILING DATE:
CLASSIFICATION:
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: 08/469,667
FILING DATE: 06-JUN-1995
ATTORNEY/AGENT INFORMATION:
TITLE OF INVENTION: Colon Specific Genes and Proteins

NUMBER OF SEQUENCES: 24
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Carella, Byrne, Bain, Gilfillian, Cecchi,
 STREET: Stewart & Olstein
 CITY: Becker Farm Road
 STATE: NJ
 COUNTRY: USA
 ZIP: 07139

COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 OPERATING SYSTEM: IBM PC compatible
 SOFTWARE: PC-DOS/MS-DOS
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: PCT/US95/07289
 FILING DATE: 06-JUN-1995
 CLASSIFICATION:
 ATTORNEY/AGENT INFORMATION:
 NAME: Ferraro, Gregory D.
 REGISTRATION NUMBER: 36,134
 REFERENCE/DOCKET NUMBER: 325800-265
 TELECOMMUNICATION INFORMATION:
 TELEFAX: 201-994-1744
 INFORMATION FOR SEQ ID NO: 17:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 605 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: cDNA
 FEATURE:
 NAME/KEY: CDS
 LOCATION: 1..603
 FEATURE:
 NAME/KEY: mat_peptide
 LOCATION: 1..603
 PCT-US95-07289-17

Query Match 31.3%; Score 284.8; DB 5; Length 605;
 Best Local Similarity 67.2%; Pred. No. 3.8e-84;
 Matches 403; Conservative 0; Mismatches 197; Indels 0; Gaps 0;

Qy 128 TTGAGTTAACCAAGAATGAAATATGAACTTGTGGCTCCGCACACTTGATCAGT 187
 Db 2 TTGATTTAACCAAGTGAACCAACATGACACCCTCTGAACACCTATTAGTGTCTCT 61
 Qy 188 ATGACCCAAAGCTAGCTAAATCATCGAAAGAAATATTCATGAACTTGTGACTTGG 247
 Db 62 ACAACCAAGCCACAGCAAGCAAGCAAGAAATATTCATGAACTTGTGAACTTGTG 121
 Qy 248 ATGACACAGAACAAATCAAGTGTGGCTCCGCACAGCTTGTGGCTCCGCAC 307
 Db 122 AGGACACAGATAACCGATTCAGTCAGTGTGGCAAGTACAATGACATACAGCT 181
 Qy 308 GGAGGTGATGTGTGAGAGTGTGTCATGTGTGAACTTGATGAACTTGTGAACT 367
 Db 182 TTCAGTCATTTCACTGGCAGTACAATGACATGGTGTGAACTTGTGAACT 241
 Qy 368 GACTGAGGTATGTGTGAGAGTGTGTCATGTGTGAACTTGATGAACTTGTGAACT 427
 Db 242 GACTGAGGTATGTGTGAGAGTGTGTCATGTGTGAACTTGATGAACTTGTGAACT 301
 Qy 428 TTGAACTTAACTCCAACTGCAAGATTAAGTGTGAGACTTGATGAACTTGTGAACT 487
 Db 302 TTGCTGAGCTGCTCAAAAGGTGATGTTGTGAAAGTGTGAAAGTTG 361
 Qy 488 GTGAACTTAACTCCAACTGCAAGATTAAGTGTGAGACTTGATGAACTTGTGAACT 547
 Db 362 GTGAGGCCAACCAAAGTGTGCAAGAATCTGATGTCATGAACTTGTGAACT 421
 Qy 548 GTAAACAAACTGATTCAAAATTGACTTAACTGCAATCTGCTTCACCATCTGGGACT 607

RESULT 4
 US-08-927-128-3
 Sequence 3, Application US/08927128
 Patent No. 6127150
 GENERAL INFORMATION:
 APPLICANT: Coolidge, Thomas
 APPLICANT: Wagner, Fred
 APPLICANT: van Heke, Gino
 APPLICANT: Schuster, Sheldon
 APPLICANT: Scout, Jay
 APPLICANT: Wylie, Dwane
 TITLE OF INVENTION: PURIFICATION DIRECTED CLOSING OF PEPTIDES
 NUMBER OF SEQUENCES: 28
 CORRESPONDENCE ADDRESS:
 ADDRESSEEE: Merchant, 5 Gould
 STREET: 3100 No. 6127150west Center, 90 S. 7th Street
 CITY: Minneapolis
 STATE: MN
 COUNTRY: U.S.A.
 ZIP: 55402
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Diskette
 COMPUTER: IBM Compatible
 OPERATING SYSTEM: DOS
 SOFTWARE: FastSEQ Version 1.5
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/927,128
 FILING DATE: 05-SEP-1997
 CLASSIFICATION: 435
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 08/680,004
 FILING DATE: 15-JUL-1995
 ATTORNEY/AGENT INFORMATION:
 NAME: Carter, Charles G
 REGISTRATION NUMBER: 35,093
 REFERENCE/DOCKET NUMBER: 8648-2USD1
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 612/332-5300
 TELEFAX: 612/332-9081
 TELEX:
 INFORMATION FOR SEQ ID NO: 3:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 780 base pairs
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: Genomic DNA
 HYPOTHETICAL: NO
 ANTI-SENSE: NO
 FRAGMENT TYPE:
 ORIGINAL SOURCE:
 FEATURE:
 NAME/KEY: Coding Sequence
 LOCATION: 1..780
 OTHER INFORMATION:
 US-08-927-128-3

Query Match Score 280; DB 3; Length 780;
 Best Local Similarity 63.4%; Pred. No. 1.7e-32;
 Matches 255; Indels 3; Gaps 1;

RESULT 5
US-08-595-868C-11
Sequence 11, Application US/08595868C

GENERAL INFORMATION:
 APPLICANT: Wagner, Fred
 APPLICANT: Stout, Jay Dennis
 APPLICANT: Henriksen, Dennis
 APPLICANT: Partridge, Bruce
 APPLICANT: Holmquist, Bartt
 APPLICANT: Frank, Julie
 TITLE OF INVENTION: RECOMBINANT PREPARATION OF CALCITONIN FRAGMENTS AND US
 NUMBER OF SEQUENCES: 51
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Merchant & Gould
 STREET: 3100 No. 5962270 West Center, 90 S. 7th Street
 CITY: Minneapolis
 STATE: MN
 COUNTRY: U.S.A.
 ZIP: 55402
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Diskette
 COMPUTER: IBM Compatible
 OPERATING SYSTEM: DOS

Qy 48 CTGGGATAACGGGAGCAACGGTCTATTCACTGGAGGAATTTCCCTATTGCTGA 107
 Db 12 CTGGGGTAGGAAACAAACGGACTGGATAAGCTCCATTGCGAA 71
 Qy 108 TGGTGATCAGCAATCTCCAAATGGATTAACAAAGGAAATATGACTCTCCCT 167
 Db 72 GGAGAGGCCAGTCCCTGTGACATGACAACTCATACGCCAAGTCCTCCCT 131
 Qy 168 CCCACCACTAGTATCAGTAACTGAGCTAGCTAACAAACAGGGCA 227
 Db 132 GAAGCCCTGTCGTTCTATGATCAAGCAACTTCCCTGAGGATCCATAGTGCNA 191
 Qy 228 TTCTTCATATGTTGACTTGTGATGACAGAAACAATCAAGTCTGGTCTGCTCT 287
 Db 192 TGTTTCGACGTTGAGTTGATGACTCAGGCAAACGGTCTCAGGGAGACCT 251
 Qy 288 CACTGGAGCTACAGTTACGGGAGTTACCTTCACGGGGTCCGGTGTGATGACAGG 347
 Db 252 GGATGGCTTAACTGATGTTGATGTTGACCTGGGGTCACTGTGACAGGG 311
 Qy 348 CTCGAGCACATAGTGGAGTGGAGCTGAGGCTCAGTGTGTCACGAA 407
 Db 312 TTCAAGGAGATACTGTGGTAAAGAAATATGTCGACAACTTCACTGGTCA 371
 Qy 408 TTCAAGACAAATACCCAGTTGGAAAGCTGGAGGAGCTGAGCTGGCTGTT 467
 Db 372 CAC - GAAAATGGGATTGGAAAGCTGGAGGAGCTGAGCTGGCTGTT 428
 Qy 468 GGGAGTTGTTTACAGATTTGGAAACCTTAATTCCCAACTGCAAAAGATTACTGACACTT 527
 Db 429 AGGTATTGTTTGGGTTGAGGTGGACGCTGAAACGGGCTTCAGAAAGTGTGATGCT 488
 Qy 528 GGATTCCATTAAAGAAGGTTAACAAACTGATTCACAAATTGACCTATTGCTCT 587
 Db 489 GGATTCCATTAAACAAAGGCAAGAGCTGCTGACTCTGACTAACTTCGATCTGGCT 548
 Qy 588 GCTTCCACATCTGGGACTACTGGACATACTGGATATCTGGTTCTCTACAGTCCACCTCTCT 647
 Db 549 CCTTCCTGAACTCTGGGAAATCTGGGACTTCTGGACCTACCCAGGCTACAGACCCPCTCT 608
 Qy 648 TGAGAGTGCACATGGATTGTTAACGGAACTTAAACATGCTCAACAGCTGC 707
 Db 669 GGATGTGTCACCTGGATGCTGGTCAAGGAACCATCTGGGTAGACGGAGGGTT 668
 Qy 708 CAATTTCGACGTTCTGGTGCACAGGGAGGTGAAGCAGAG 751
 Db 669 GAAATTCCATTAACTTACTCATGGGAGGTGAACCCGAG 712
 Qy Sequence 11, Application US/08595868C
 Patent No. 5962270
 GENERAL INFORMATION:
 APPLICANT: Wagner, Fred
 APPLICANT: Stout, Jay Dennis
 APPLICANT: Henriksen, Dennis
 APPLICANT: Partridge, Bruce
 APPLICANT: Holmquist, Bartt
 APPLICANT: Frank, Julie
 TITLE OF INVENTION: RECOMBINANT PREPARATION OF CALCITONIN FRAGMENTS AND US
 NUMBER OF SEQUENCES: 51
 CORRESPONDENCE ADDRESS:
 ADDRESSSEE: Merchant & Gould
 STREET: 3100 No. 5962270 West Center, 90 S. 7th Street
 CITY: Minneapolis
 STATE: MN
 COUNTRY: U.S.A.
 ZIP: 55402
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Diskette
 COMPUTER: IBM Compatible
 OPERATING SYSTEM: DOS

SOFTWARE: FastSEQ Version 1.5
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/595, 868C
 FILING DATE: 06-FEB-1996
 CLASSIFICATION: 435
 PRIOR APPLICATION NUMBER:
 APPLICATION NUMBER:
 FILING DATE:
 ATTORNEY/AGENT INFORMATION:
 NAME: Carter, Charles G
 REGISTRATION NUMBER: 35,093
 REFERENCE/DOCKET NUMBER: 8648 . 59US01
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 612/332-5300
 FAX: 612/332-9081
 TELEX:
 INFORMATION FOR SEQ ID NO: 11:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 864 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: Genomic DNA
 HYPOTHETICAL: NO
 ANTI-SENSE: NO
 FRAGMENT TYPE:
 ORIGINAL SOURCE:
 FEATURE:
 NAME/KEY: Coding Sequence
 LOCATION: 1...864
 OTHER INFORMATION:
 US-08-595-868C-11
 Query Match Similarity 30.7%; Score 280; DB 2;
 Best Local Similarity 63.4%; Pred. No. 1..8..82;
 Matches 446; Conservative 0; Mismatches 255; Indels 3; Gaps 1;
 Db 48 CTGGGATACCGGGAGCAACGGCTTATTCACTGGAGGAAATTCCCTPATGCTGA 107
 Db 12 CTGGGTACGGAAACAAAGGACCTGAGCTGACTGGATTCATGGCAA 71
 Qy 48 TGTTGATGACAAATCTCCAATGGATTAACCAAAAGGAAATATGACTCTTCCT 167
 Db 72 GGAGAGCCACTCCCTGTCTCTGATCAAGCAACTCATGACAGTACATGGCT 131
 Qy 168 CCCACCACTTGTGACTTCAAGTGAACCTGACCTAAATCTGAAACAGGGCA 227
 Db 132 GAAGGCCCTGTCCTCTGATCAAGCAACTCCCTGAGGATCCCTCAAAATGCT 191
 Db 228 TTCTTCATATGTTGACTTGTGACTTGTGACTTGTGACTTGTGACTTGTGACTT 287
 Db 192 TGTTTACAGTGGAGTTGATGACTCTGAGCTCAGGAAACGGGAGCCCT 251
 Qy 288 CACTGGACACATAGTGGATATGGTGTGACTCTGGGTTCCGCTGATGACAGG 347
 Db 312 TTCAAGGACATACTGTGGATAAAAGAAATATCTGCAAACTTCACTGGTCACTGGTCA 371
 Qy 408 TTCAAGACAAATCTCCAATGGATTAACCAAAAGGAAATATGACTCTTCCT 167
 Db 72 GGAGAGCCACTCCCTGTCTCTGATCAAGCAACTCATGACAGTACATGGCT 131
 Qy 468 CCCACCACTTGTGACTTCAAGTGAACCTGACCTAAATCTGAAACAGGGCA 227
 Db 132 GAAGGCCCTGTCCTCTGATCAAGCAACTCCCTGAGGATCCCTCAAAATGCT 191
 Qy Sequence 11, Application US/08595868C
 Patent No. 5962270
 GENERAL INFORMATION:
 APPLICANT: Wagner, Fred
 APPLICANT: Stout, Jay Dennis
 APPLICANT: Henriksen, Dennis
 APPLICANT: Partridge, Bruce
 APPLICANT: Holmquist, Bartt
 APPLICANT: Frank, Julie
 TITLE OF INVENTION: RECOMBINANT PREPARATION OF CALCITONIN FRAGMENTS AND US
 NUMBER OF SEQUENCES: 51
 CORRESPONDENCE ADDRESS:
 ADDRESSSEE: Merchant & Gould
 STREET: 3100 No. 5962270 West Center, 90 S. 7th Street
 CITY: Minneapolis
 STATE: MN
 COUNTRY: U.S.A.
 ZIP: 55402
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Diskette
 COMPUTER: IBM Compatible
 OPERATING SYSTEM: DOS

RESULT 6

US-09-139-819A-11
Sequence 11, Application US/09139819A
Patent No. 6251635

GENERAL INFORMATION:

APPLICANT: WAGNER, Fred W.
STOUT, Jay S.

APPLICANT: HENRIKSEN, Dennis B.
PARTRIDGE, Bruce E.

APPLICANT: HOLMOUST, Bart

APPLICANT: FRANK, Julie A.

TITLE OF INVENTION: RECOMBINANT PREPARATION OF CALCITONIN
TITLE OF INVENTION: FRAGMENTS AND USE THEREOF IN THE PREPARATION OF CALCITONIN
TITLE OF INVENTION: AND RELATED ANALOGS
NUMBER OF SEQUENCES: 51

CORRESPONDENCE ADDRESS:
ADDRESSEE: FOLEY & LARDNER
STREET: 3000 K Street, N.W.
CITY: Washington
STATE: D.C.
COUNTRY: U.S.A.
ZIP: 20007-5109

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/139, 819A
FILING DATE: 25-AUG-1998
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/595, 868
FILING DATE: 06-FEB-1996

ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29, 768
REFERENCE/DOCKET NUMBER: 089187/0144

TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 672-5300
TELEFAX: (202) 672-5399

SEQUENCE CHARACTERISTICS:
LENGTH: 867 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)

FEATURE:
NAME/KEY: CDS
LOCATION: 1..864

US-09-139-819A-11

Qy 568 GCTTCCACCACCTGGGACTACTGGGATATCCTGGTTCTTACAGTTCCACCTTCCT 647
Db 549 CCTTCCTGATCCTGGATTCCTGGATTAACGGCTACCCGCTACTGACCACTCCCTCT 608

Qy 648 TGAGATGTCACTGGATTGTTAAAGCAACCTATAAACATCGTCTAACAGCTGGC 707
Db 609 GAAATGTTGACCTGGATTGCTCAAGAACCCATCAGGTCAGAGGAGCTGTT 668

Qy 708 CAAATTTCGCACTCTCTGGAACGGAGGGTGAGGAGCG 751
Db 669 GAATTCGTAACCTTAACCTTAATGGGAGGGTGAACCGAAG 712

Qy 288 CACTGGAAAGCTACAGGTTACCGCAGGTTACCTTCAGTGGGTCGGCTGATGCCACGG 347
Db 252 GGATGGACATTCAGATGATTCAGTTGATTCAGTTGAGTGGAAAGG 311

Qy 348 CTCCGAGCACATAGTGGATGGAGCTATGCTGAGAGCTCCATGTTGGTCACGGAA 407
Db 312 TTCAAGGCATACTGTCGATATAAAAGAAATGCTGAGAACTTCACCTGGTCACTGGAA 371

Qy 408 TTCAGAAATAACCCAGCTTGTAGGAGCTCTGAAACAGATGGACTGGCTTCTT 467
Db 372 CAC--CAAATATGGGATTITGGGAAAGCTGTCGCAACTGTGACTGGCGTCT 428

Qy 468 GGGAGTGTCTTTACGATTGGTAACCTAAATCCAAACTGCAAAAGATACTGACACTT 527
Db 429 AGGTATTTTGAAAGTTGGAGGGTAACCGGGCTTCAAGAATGTTGATGTGCT 488

Qy 528 GGATTCCATTAAAGAAAAGGTTAAACAAACTCGATICAACATTGACCTATTGTCCTCT 587
Db 489 GGATTCCATTAAACAAAGGGCAAGTAGTGGCTGACTCTACTACTCTCGTGCCT 548

Qy 588 GCTTCACCATCTGGACTACTGGATATCTGGTTCTTACAGTTCACCTCTCT 647
Db 549 CCTTCCTGAACTCTGGATTACTGGCTTACCTGGCTTACCTGACCACTCCCTCTCT 608

Qy 648 TGGAGGTGTCACTGGATTGGTTAAAGCAACCTAAACATCAGGCTCTCAACAGCTGGC 707
Db 609 GGATGTGTGACTGGATTGGCTCAAGGAACCTAGCTGTCAGCGGAGGTGTT 668

Qy 708 CAAATTTCGCACTCTGGTCAAGGAGGGTGAGGAGCG 751
Db 669 GAATTCGTAACCTTAATGGGAGGGTGAACCCGAG 712

RESULT 7

US-09-750-913-11
Sequence 11, Application US/09750913
Patent No. 6410707

GENERAL INFORMATION:
APPLICANT: WAGNER, Fred W.
STOUT, Jay S.
HENRIKSEN, Dennis B.
PARTRIDGE, Bruce E.
HOLMQUIST, Bart
FRANK, Julie A.

TITLE OF INVENTION: RECOMBINANT PREPARATION OF CALCITONIN
FRAGMENTS AND USE THEREOF IN THE PREPARATION OF CALCITONIN
AND RELATED ANALOGS

NUMBER OF SEQUENCES: 51

CORRESPONDENCE ADDRESS:
ADDRESSEES: FOLEY & LARDNER
STREET: 3000 K Street, N.W.
CITY: Washington
STATE: D.C.
COUNTRY: U.S.A.
ZIP: 20007-5109

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/139, 819A
FILING DATE: 25-AUG-1998
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/595, 868
FILING DATE: 06-FEB-1996

ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29, 768
REFERENCE/DOCKET NUMBER: 089187/0144

TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 672-5300
TELEFAX: (202) 672-5399

SEQUENCE CHARACTERISTICS:
LENGTH: 867 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)

FEATURE:
NAME/KEY: CDS
LOCATION: 1..864

Query Match Similarity 30.7%; Score 280; DB 3; Length 867;
Best Local Similarity 63.4%; Pred. No. 1 8e-82;
Matches 446; Conservative 0; Mismatches 255; Indels 3; Gars 1;

Qy 48 CTGGGGATACCGGAGAACGGTCTATTACTGGAGGAATTTCCTTATGGCTGA 107
Db 12 CTGGGGATACGGCAAAACGACCTGAGCTAAAGGACTCCCATGGCA 71

CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09/750,913
 FILING DATE: 12-Jan-2001
 CLASSIFICATION: <Unknown>
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 09/139,819
 FILING DATE: <Unknown>
 ATTORNEY/AGENT INFORMATION:
 NAME: BENT, Stephen A.
 REGISTRATION NUMBER: 29,768
 REFERENCE/DOCKET NUMBER: 089187/0144
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (202) 672-5300
 TELEFAX: (202) 672-5399
 INFORMATION FOR SEQ ID NO: 11:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 867 base pairs
 TYPE: nucleic acid
 STRANDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: DNA (Genomic)
 FEATURE:
 NAME/KEY: CDS
 LOCATION: 1..864
 SEQUENCE DESCRIPTION: SEQ ID NO: 11:
 ; US-09-750-913-11

Query Match 30.7%; Score 280; DB 4; Length 867;
 Best Local Similarity 63.4%; Pred. No. 1..8e-82;
 Matches 446; Conservative 0; Mismatches 255; Indels 3; Gaps 1;

Qy 48 CTGGGATACCGGAGGACAACCGTCTATTCACTGGAAATTTCCTATTGCTGA 107
 Db 12 CTGGGGGAGCGGAAACACAACGGACCTGCACTGGCATTAAGGACTCCATGCCAA 71

Qy 108 TGTGTGATAGGCATCTCDAATTGGATAAACAGAGTGAAATACTCTTCCT 167
 Db 72 CGGAGAGGGCCAGTCCTCTGTGACATCGACACTCATAGCCAAAGTATGCCAA 131

Qy 168 CGGACCACTTAGTATCAAGTGAACCCAAATCATGCCAAAGCTCACTGGCC 227
 Db 132 GAAGCCCCCTGTCTGTCTCTATGATCAAGAACCTCCAGGATCTCACATGGCA 191

Qy 228 TTCCCTTCATGTGACTTTGATGACACAGAAACAATTCAGTTCTGGCTGTCCT 287
 Db 192 TGTTTCACGTGAGTGTGATGACTCTAGGAACAGTGTCAAGGAGGACCC 251

Qy 288 CACTGGAGAGCTACAGGTTACGGGAGGTACCCCTCACTGGGTCGGCTCATGCCG 347
 Db 252 GGATGGCCTTACAGATTGATTCAGTTTCACTGGGTCACCTCATGGCAAGA 311

Qy 348 CTCCGAGACATAGTAGTGGAGTGGATGCTCATGTTCACTGGAA 407
 Db 3112 TICAGAGGATACGTGGATAAAAGAAAATGTGCAAGACTTCATCTGGCTGA 371

Qy 408 TTGAGACAAATACCCAGTTTACAGATGGTGAACCTAACTCCAAAGATACTGACACTT 527
 Db 372 CAC--CTAAATGGGATTGGAAAGCTGTGCAAGCCTGTGCAAGTGTGGCTCT 428

Qy 468 GGGAGGTGTTTACAGATGGTGAACCTAACTCCAAAGATACTGACACTT 527
 Db 429 AGTATTGTTGAGGTGGCAAGGCTAACAAACTGATGGTGAACCTTGTGCT 488

Qy 528 GGATTCCATTAAAGAAAGGTTAACAAACTGATGGTCAAAATTGACCTTGTCT 587
 Db 489 GGATTCCATTAAACAAAGGCTAACAACTGATGGTCAAAATTGACCTTGTCT 548

Qy 588 GGTTCACATCTGGGACTACTGGACATCCGGTCTCTACAGTCACCTCTCT 647
 Db 549 CCTTCCTGAACTCTTGGATTACGGACTACCCAGGCTACTGACCACTCTCT 608

Qy 648 TGAGAGTCACATGGATTTGTTAAGGCAACCTAAACATCAGCTCAACACTGGC 707
 ; RESULT 8
 US-09-976-594-643
 ; Sequence 643, Application US/09976594
 ; Patent No. 6673549
 ; GENERAL INFORMATION:
 ; APPLICANT: Furness, Michael
 ; APPLICANT: Buchbinder, Jenny
 ; TITLE OF INVENTION: GENES EXPRESSED IN C3A LIVER CELL CULTURES TREATED WITH STEROIDS
 ; FILE REFERENCE: PA-0041 US
 ; CURRENT APPLICATION NUMBER: US/09/976,594
 ; CURRENT FILING DATE: 2001-10-12
 ; PRIORITY APPLICATION NUMBER: 60/10240,409
 ; PRIORITY FILING DATE: 2000-10-12
 ; NUMBER OF SEQ ID NOS: 1143
 ; SOFTWARE: PERL, Program
 ; SEQ ID NO: 643
 ; LENGTH: 1759
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; NAME/KEY: misc feature
 ; OTHER INFORMATION: Invryte ID No. 6673549 2742913CB1
 ; US-09-976-594-643

Query Match 30.7%; Score 280; DB 4; Length 1759;
 Best Local Similarity 63.4%; Pred. No. 2.e-82;
 Matches 446; Conservative 0; Mismatches 255; Indels 3; Gaps 1;

Qy 48 CTGGGATACCGGAGGACAACCGTCTATTCACTGGAAATTTCCTATTGCTGA 107
 Db 273 CTGGGGTACGGGAAACACAACGGACCTGAGCTGGCATTCATGCCAA 332

Qy 108 TGTTGATAGGCATCTCDAATTGGATAAACAGAGTGAAATACTGCTCTCC 167
 Db 333 CGGAGAGGGCCAGTCCTCTGTGTCACATGCCAACTGACACTCATGCCA 392

Qy 168 CGGACCACTTAGTATCAAGTGAACCCAAATCATGCCAAAGCTCACTGGCC 227
 Db 393 GAAGCCCCCTGTCTGTCTCTATGATCAAGAACCTCCAGGATCTCACATGGCA 452

Qy 228 TTCTCTCAATGTTGACTTTGATGACACAGAAACAATTCAGTTCTGGCTGTCCT 287
 Db 453 TCTTTCAACGGTGGATTGATGACTCTCAGGCAAAGGAGTGTCAAGGGAGGACCCCT 512

Qy 288 CACTGGAGACATAGTAGTGGAGTGGATGCTCATGTTCACTGGGGTCGGTGTGATGCCAGG 347
 Db 513 GGATGGCCTTACAGATGATGTCATCTGATGGCTCATGTTCACTGGGGTCACTGATGCCAGG 572

Qy 348 CTCCGAGACATAGTAGTGGAGTGGATGCTCATGTTCACTGGGGTCGGTGTGATGCCAGG 407
 Db 573 TICAGAGGATACGTGGATAAAAGAAAATGTGCAAGACTTCATCTGATGGCTCT 632

Qy 408 TTGAGACAAATACCCAGTTTACAGATGGTGAACCTAACTCCAAAGATACTGACACTT 527
 Db 633 CAC--CTAAATGGGATTGGAAAGCTGTGCAAGACCTGATGGCTCT 689

Qy 468 GGGAGGTGTTTACAGATGGTCAAACTTACTCCAAAGTGTCTCT 527
 Db 690 AGGTATTGTTGAGGTGGCAAGGCTTCAAGAAAGTGTGATGTGCTCT 749

Qy 528 GGATTCCATTAAAGAAAGGTTAACAAACTGATGGTCAAAATTGACCTTGTCTCT 587
 Db 750 GGATTCCATTAAACAAAGGCTAACAACTGATGGTCAAAATTGACCTTGTCTCT 809

Qy 588 GCTCCACCATCTGGGACTACTGGATTTGTTAAGGCAACCTAAACATCAGCTCAACACTGGC 707

| | | |
|--|---|-----|
| 810 | CCCTCCTGTAATCCTGGATTAATGGACCTTACTGGGCTCACTGACCACTCCCTCTCTCT | 869 |
| 648 | TGAGACTGTACATGGATTGTTAAAGCACCTATAAACATCGTCACAGCTGC | 707 |
| 870 | GGAAATGGTGAACCTGTGATTGCTCAAGAACCCATCGCTGAGGCCAGGTGT | 929 |
| 708 | CAAATTTCGAGTCCTGTGACAAGGGGGTAGGAGCGAG | 751 |
| 930 | GAAATTCGGTAACCTAACTTCAATGGAGGGTAGACCAGAG | 973 |
| SEQUENCE 9 | | |
| Sequence 124, Application US/09566921 Patent No. 6628889 | | |
| GENERAL INFORMATION: | | |
| APPLICANT: Loring, Jeanne F. | | |
| APPLICANT: Tingley, Debora W. | | |
| APPLICANT: Edwards, Carla M. | | |
| TITLE OF INVENTION: GENES EXPRESSED IN ALZHEIMER'S DISEASE | | |
| FILE REFERENCE: PA-0024 US | | |
| CURRENT APPLICATION NUMBER: US/09/566,921 | | |
| NUMBER OF SEQ ID NOS: 138 | | |
| SOFTWARE: PERL Program | | |
| SEQ ID NO: 124 | | |
| LENGTH: 1755 | | |
| TYPE: DNA | | |
| ORGANISM: Homo sapiens | | |
| FEATURE: | | |
| NAME/KEY: misc_feature | | |
| OTHER INFORMATION: Incyte ID No. 6682888 248306.1 | | |
| LOCATION: 1751-1752 | | |
| OTHER INFORMATION: a, t, c, g, or other | | |
| SEQUENCE 124 | | |
| Query Match Score 30.6%; DB 4; Length 1755; | | |
| Best Local Similarity 63.2%; Pred. No. 9-7e-82; | | |
| Matches 445; Conservative 0; Mismatches 256; Indels 3; Gaps 1; | | |
| 48 CTGGGGATACCCGAGAACCGTCTTACCTGGAGGAATTTCCTTATTTGCTGA 107 | | |
| 273 CTGGGGTAAACAAACGACTCTGGGAACTGAGCTGGATAAGACTCCCATGGCA 332 | | |
| 108 TGGTGAATCGCAATCTCAATGGATAAAAGAATGAAATATGACCTTCCT 167 | | |
| 333 GCGAGAGCCAGTCCTGGAACTGAGCTCATGGAGATGACCTTCCT 392 | | |
| 168 CGACCAACTAGTATCAAGTATGACCCAGACTAGCTAACATCGGAACGGCA 227 | | |
| 393 GAAGCCCCCTGTCTGTTCTCATGTCAGAACATTCCTGGATCCIAACATGGCA 452 | | |
| 228 TTCCCTCAATGTGACTTGTGACCACTGGAGAACAAATGAGTCTGGGGTGTCT 287 | | |
| 453 TGTTTCACGGGAGTTGATCACTTGAGATAAGGAGACCCCT 512 | | |
| 288 CACTGGAAAGCTACAGGTGACGGTCACTCTACGGGGTCACTGGTGTGACCAAGG 347 | | |
| 513 GGATGGCAGCTTACAGATGATTGAGCTGATCTTCACTGGGTTCACITGTGACAAAGG 572 | | |
| 348 CTCGGAGCAACATAGTGGAGTGGACTGAGCTGAGCTGTCATGTGAA 407 | | |
| 573 TTCAAGGATTAATGTGGATAAAAAGAAATATGCTGCAAGACTCACTGGTCACTGGAA 632 | | |
| 408 TTGAGACAAATAACCCAGCTTGTGAGCCAGCTATGAGCTGACTGGGTCTCT 467 | | |
| 633 CAC--CAATATGGGATTTGGAAAGGTGAGCAACCTGATGGATGGGTCT 689 | | |
| 468 GGAGATGTGTTTACAGATGGTGAACCTAATTCCAACCTGAAAGATGACACTT 527 | | |
| 690 AGGTATTGTTTGAGGTTTACAGATGGTGAACCTAATTCCAACCTGAAAGATGACACTT 749 | | |

RESULT 1.1
US-09-385-982-446

Sequence 446, Application US/09385982

PATENT NO. 626234

GENERAL INFORMATION:

APPLICANT: ENDEGE, WILSON O., ET AL.

TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION

FILE REFERENCE: CCDNA-260XX

CURRENT FILING DATE: 1999-08-30

EARLIER APPLICATION NUMBER: 09/328,111

EARLIER FILING DATE: 1999-06-08

EARLIER APPLICATION NUMBER: 60/117,393

EARLIER FILING DATE: 1999-01-27

EARLIER APPLICATION NUMBER: 60/098,639

EARLIER FILING DATE: 1998-08-31

NUMBER OF SEQ ID NOS: 544

SOFTWARE: FastSEQ For Windows Version 3.0

SEQ ID NO: 446

LENGTH: 599

TYPE: DNA

ORGANISM: Homo sapiens

FEATURE: misc_feature

NAME/KEY: misc_feature

LOCATION: (1)..(59)

OTHER INFORMATION: n = A,T,C or G

US-09-385-982-446

Query Match 19.5%; Score 177.4; DB 3; Length 599;

Best Local Similarity 61.0%; Pred. No. 1..e-48;

Matches 351; Conservative 0; Mismatches 217; Indels 9; Gaps 4;

QY 53 GATACCGGGGACAACGGTCTATTCACTGGAGGAATTTCCTTAATTGTGCTG 112

DB 1 GGACGGAAACACAACGACCTGAGCCTGGCATTCGGCAAGGGAG 60

QY 113 ATCGCAATCTCCAAATTGAGATAAACAAAGTAAATGACTTCCTCCAC 172

DB 61 AGGCCAGTCCTCTGTTGACATGACATCACAGCAAATGACCTCCCTGAAG 120

QY 173 CACTTAGTATCAAGTATGACCCAAAGTCAGCTAAATCATCGAACGGGCATTCC 232

DB 121 CCCCTGCTGTTCTCATGTCATGCAACTCCCTGAGATCCTAACATGGTATGTT 180

QY 233 TCAATGTTGACTTGTGACACAGAGAACAAATCAGTTCCTGGCTCCTCACTG 292

DB 181 TCAACGTGGAGTTGATGACTTCAGGCAAAGTGTCAAGGGAGAACCCCTGGATG 240

QY 293 GAAGCTATGGTTACGGGAGTTCACTTCAGCTGGGTCTCGTATGACAGGCTCG 352

DB 241 GCACCTTACAGTGGATGACTTCAGTGGGTCACTTGATGACAAAGTTCACT 300

QY 353 AGACATAGTGTGAGATGGAGTCACTTCAGCTGGGTCTCGTATGAGTCAGTGGAAATTAG 412

Db 913 GGATTTCCTGAGATAGGACATGAGAATGGGAGITCCAGATTTCTTGATGATGTG 972

QY 529 GATTCCTTAAAGAAAAGGTTAACAAACTCGATTCAAATTGAGCTATGTCCTGTG 598

Db 973 GACAGATTAAGAACGGCAAGGGCCCTTCAGAAAGTTGACCCATCTGCTG 1032

QY 589 CTTCACCATCTGGACTACTGACATATCCCTGTTCTTCAGTTCACCTCTT 648

Db 1033 TTCCGGATGCCGGACTACTSGACCAACGGCTTATTGCCACCCGCTGGAG 1092

QY 649 GAGGTGTACATGATTAAAGCAACCTATAAACATCAAGCTCTAACAGCTGCC 708

Db 1093 GAATGATCTGTGTCGTCAGGCCCCATGACGTGACCATGCCATGCC 1152

QY 709 AAATTGCGAGTCCTGCTGATAGCGAGGGTAAAGCAGCAGTCTTCT 758

Db 1153 AAGCTGGAGCCCTCTCCAGTGTGCTGAGAACGCCAGTNGCCCT 1202

Db 301 AGCTACTGTGAAAGAATAATGCTGAGAACCTCACTGGTACTGAAACAC- 358

QY 413 ACAATACCCGGTTTGTGAGCGCTCATGACATGGCTGACTGCTG-TCTTGGA 471

Db 359 -CAATAATGGGATTGGAAGCTGTCAGGAAACCTGATGCACTGCGTTCTAGTGA 417

QY 472 GTGTTTACAGATGGTGAACCTAATTCCAACACTGCAAAGATACTGACACT-TTCTGA 530

Db 418 TTTTGAAGGGTGGAGCTAAACNGGCTTNTAAACTGTGAATGTCGTTGA 477

QY 531 TCCATTAAAGAAAGGTTAACAAACTCGATTCACAAATTGACCTATTGCTCTGCT 590

Db 478 TTCCATTAAACAAAGGCCAGATTGCTGACTTCACTAATTNNAAATCCTCTNGGCT 537

QY 591 TCCACC---ATCCTGGACTATGGACATATCTGTT 625

Db 538 TCTTCCTGAAATCCTGATTACGGACTNCAGCT 576

RESULT 1.2
US-09-621-976-18059

Sequence 18059, Application US/09621976

PATENT NO. 6639063

GENERAL INFORMATION:

APPLICANT: Dumas Mline Edwards, J.B.

APPLICANT: Robert S.

APPLICANT: Giordano, J.Y.

TITLE OF INVENTION: ESTS and Encoded Human Proteins.

FILE REFERENCE: GENSET_054PR2

CURRENT APPLICATION NUMBER: US/09/621,976

CURRENT FILING DATE: 2000-07-21

NUMBER OF SEQ ID NOS: 19335

SEQ ID NO: 18059

SOFTWARE: Patent.pml

LENGTH: 477

TYPE: DNA

ORGANISM: Homo sapiens

US-09-621-976-18059

Query Match 16.2%; Score 147.6; DB 4; Length 477;

Best Local Similarity 65.6%; Pred. No. 9..e-39;

Matches 244; Conservative 1; Mismatches 125; Indels 2; Gaps 2;

QY 51 GGGATACCGGAGACAAAGGTCTATTICACTGGAGAAATTTCCTTATTCGATGG 110

Db 66 GGCGCTAGCAGTCACAGGTCTGACACTGGCATGAACTTCCAAATGCCAGGG 125

QY 111 TGATCGAAATCTCAATTGAGATAAACCAAAGAAGTGAATAATGACTCTCCCTCG 170

Db 126 GGAAACCAAGTCGCCRTGAGCTGAACTAGATGACATGAGATGAGTC 185

QY 171 ACCACTTAGTATCAAGTATGACCAAGTTCCTTAATCATCGAACAGGGCCATTG 230

Db 186 GCGATGGCTCTGCTCTCTATGATGGCTCTGCAATGGAAAGAC 244

QY 231 CTCATGTTGACTTTGATGACACAGAAACAAATCAAGTTCTGCTCTAC 290

Db 245 CTGGCGAGTTGATTTGATGATGAGTCATGGTCAATGCTGAGATGATGGTC 304

QY 291 TGGAGGTACAGGTACGGCTTACCTTCAGTGGCTGCTGATGACAGGGCTC 350

Db 305 TGAACCTACGGACTTCCTGCTCTGAGATCTTCATGCTGAGCTCTCTC 364

QY 351 CGAGCACATAGTGTGAGTGGAGTGGCTATGCTGAGAGCTCCATGTTGTTCACTGGATTTC 410

Db 365 TGACACACCGTGGATGAGTCAGTGGAGTCACTGGAG-TTCATTGGTTCACTGGAAACC 423

QY 411 AGCAATAACCC 422

Db 424 GAGTATAACAC 435

RESULT 13
US-09-385-982-136
Sequence 136, Application US/09385982
Patent No. 6262334
GENERAL INFORMATION:
APPLICANT: ENDEGE, WILSON O., ET AL.
TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION
TITLE OF INVENTION: PRODUCTS: II
FILE REFERENCE: CCDNA-260XX
CURRENT APPLICATION NUMBER: US/09/385,982
CURRENT FILING DATE: 1999-08-30
EARLIER APPLICATION NUMBER: 09/328,111
EARLIER FILING DATE: 1999-06-08
EARLIER APPLICATION NUMBER: 60/117,393
EARLIER FILING DATE: 1999-01-27
EARLIER APPLICATION NUMBER: 60/098,639
EARLIER FILING DATE: 1998-08-31
NUMBER OF SEQ ID NOS: 544
SOFTWARE: FastSEQ for Windows Version 3.0
SEQ ID NO 136
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1) . (581)
OTHER INFORMATION: n = A,T,C or G
US-09-385-982-136

Query Match 16.2%; Score 147.4; DB 3; Length 581;
Best Local Similarity 65.8%; Prod. No. 1..2e-38;
Matches 214; Conservative 0; Mismatches 111; Indels 0; Gaps 0;

Qy 419 ACCCAGCTTGTGAGCAAGCTCATGAACCGATGGACTGCCTTGGAGGTGTT 478
Db 1 ACTCAGCCTTGTGTAAGCTGCTCAAGGCTGATGGTTGCOAGTTATGGTTGTA 60
Qy 479 TACAGATTTGTGACCTTAATGCCAACCTGAAAGATAATTCTGACACTTGTGATTCCA 538
Db 61 TGAGGTGTGAGGCCAACCAAAGTGAGACTATGCCCCAGCAATTAA 120

RESULT 14
US-09-385-982-156/C
Sequence 156, Application US/09385982
Patent No. 6262334
GENERAL INFORMATION:
APPLICANT: ENDEGE, WILSON O., ET AL.
TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION
TITLE OF INVENTION: PRODUCTS: II
FILE REFERENCE: CCDNA-260XX
CURRENT APPLICATION NUMBER: US/09/385,982
CURRENT FILING DATE: 1999-08-30
EARLIER APPLICATION NUMBER: 09/328,111
EARLIER FILING DATE: 1999-06-08
EARLIER APPLICATION NUMBER: 60/117,393
EARLIER FILING DATE: 1999-01-27
EARLIER APPLICATION NUMBER: 60/098,639
EARLIER FILING DATE: 1998-08-31

Query Match 12.2%; Score 111; DB 3; Length 586;
Best Local Similarity 62.4%; Prod. No. 1..4e-26;
Matches 174; Conservative 0; Mismatches 105; Indels 0; Gaps 0;

Qy 519 TGCACCTGTGACAGCGAGGGTGA 743
Db 301 GCCTTCATCAAATGTTAGGTGA 325

RESULT 15
US-09-385-982-472
Sequence 472, Application US/09385982
Patent No. 6262334
GENERAL INFORMATION:
APPLICANT: ENDEGE, WILSON O., ET AL.
TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION
TITLE OF INVENTION: PRODUCTS: II
FILE REFERENCE: CDDNA-260XX
CURRENT APPLICATION NUMBER: US/09/385,982
CURRENT FILING DATE: 1999-08-30
EARLIER APPLICATION NUMBER: 09/328,111
EARLIER FILING DATE: 1999-06-08
EARLIER APPLICATION NUMBER: 60/117,393
EARLIER FILING DATE: 1999-01-27
EARLIER APPLICATION NUMBER: 60/098,639
EARLIER FILING DATE: 1998-08-31
NUMBER OF SEQ ID NOS: 544
SOFTWARE: FastSEQ for Windows Version 3.0
SEQ ID NO 472
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1) . (586)
OTHER INFORMATION: n = A,T,C or G
US-09-385-982-472

Query Match 15.0%; Score 136.8; DB 3; Length 760;
Best Local Similarity 66.0%; Prod. No. 4..8e-15;
Matches 198; Conservative 0; Mismatches 102; Indels 0; Gaps 0;

Qy 30 GACCATOTCGAGGTGAGCTGGATACGGAGACAACGSGTCCATTACTGGAGGA 89
Db 310 GATATGGCAGTCAGCTGGATGGGATATGATGACAAAATGCTGGAAATGGCAA 251
Qy 90 ATTTTCCTATTGCTGATGGTGTGACTGAACTCTCCATTGAGTAAACAAAGAGT 149
Db 250 GCTGTATCCCATGCTGCAATGGAAATACAGTCCCCTGTGATTAACAGGAAAC 191
Qy 150 GAATATGACCTCTTCCCTCCACCATTGATCAGTACGCTGAGCTGAGCTAAAT 209
Db 190 CAAACATGACACCTCTGAAACCTTATGCTCTCTACACCCGCCACAGCCAAA 131
Qy 210 CATCAGAAACAGGGCATTCCTCAATGTTGACTTGTGATCACAGAGAAACAAATG 269
Db 130 AATTTTATGTTGGGATTCCTCATGTAATTGAGGCAACATACCGATAGT 71
Qy 270 TCTGCGTGTGGTCCCTCACTGGAACTCAGGTACGGTACCTTCACTGGGG 329
Db 70 GCTGAAAGGTGCTCTGACAGCTACAGGTCTTCAAGGCTTCAAGGCT 11

| | | | |
|----|-----|---|-----|
| Db | 7 | TGATGCCCTCCAAGCAATTAAAACCAAGGGCAAACGGACCCATTCAAATTTGACCC | 66 |
| Qy | 579 | ATTTCTCGCTTCCACCATCCCTGGACTACTGGACATAATCCCTGGTCTCTAACGTTCC | 638 |
| Db | 67 | CTCTACTCTCCTCTTCCTTATCCCTGGATTTCTGACCTACCCCTGGCTCTGACTCATCC | 126 |
| Qy | 639 | ACCTCTCTTGAGAGTGACATGGATGTTAAAGCAACCTATAAACATGAGCTCTCA | 698 |
| Db | 127 | TCCCTCTATGAGGTGAACCTGGATATCTTAAGGAGAGATCACTGAGCTAGA | 186 |
| Qy | 699 | ACAGCTGCCAATTTCGAGTCCTCTGCAAGCGAGGTGAAGGAGCTTTCT | 758 |
| Db | 187 | GCAGCTGGACAATTCAAGGCCCTCATCAAAATGTTAAGGTGATAAGCCTCTCCCAT | 246 |
| Qy | 759 | GTTATAGAGTCACTCCTGACCCAGGGTGGGGAG 797 | |
| Db | 247 | GCAGCACAAACGGCCAAACCCAAACCTCTGAAGGGAG 285 | |

Search completed: September 14, 2004, 00:33:17
 Job time : 99 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 13, 2004, 22:47:51 ; Search time 499 Seconds
9185.806 Million cell updates/sec
(without alignments)

Title: US-10-069-434-4

Perfect score: 911

Sequence: 1 cggaaatccggctcgaaatggttcc.....ccccaaatgcccaggggaaaa 911

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 3304383 seqs, 2515761380 residues

Total number of hits satisfying chosen parameters: 6608766

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Published Applications NA:
 1: /cgn2_6/_ptodata/2/pubpna/us07_PUBCOMB.seq:*
 2: /cgn2_6/_ptodata/2/pubpna/pct_new_pub.seq:*
 3: /cgn2_6/_ptodata/2/pubpna/us06_NEW_PUB.seq:*
 4: /cgn2_6/_ptodata/2/pubpna/us06_PUBCOMB.seq:*
 5: /cgn2_6/_ptodata/2/pubpna/us07_NEW_PUB.seq:*
 6: /cgn2_6/_ptodata/2/pubpna/us08_PUBCOMB.seq:*
 7: /cgn2_6/_ptodata/2/pubpna/us08_PUBCOMB.seq:*
 8: /cgn2_6/_ptodata/2/pubpna/us08_PUBCOMB.seq:*
 9: /cgn2_6/_ptodata/2/pubpna/us09B_PUBCOMB.seq:*
 10: /cgn2_6/_ptodata/2/pubpna/us09C_PUBCOMB.seq:*
 11: /cgn2_6/_ptodata/2/pubpna/us09_NEW_PUB.seq:*
 12: /cgn2_6/_ptodata/2/pubpna/us09_NEW_PUB.seq:*
 13: /cgn2_6/_ptodata/2/pubpna/us10A_PUBCOMB.seq:*
 14: /cgn2_6/_ptodata/2/pubpna/us10A_PUBCOMB.seq:*
 15: /cgn2_6/_ptodata/2/pubpna/us10B_PUBCOMB.seq:*
 16: /cgn2_6/_ptodata/2/pubpna/us10C_PUBCOMB.seq:*
 17: /cgn2_6/_ptodata/2/pubpna/us10_NEW_PUB.seq:*
 18: /cgn2_6/_ptodata/2/pubpna/us60_NEW_PUB.seq:*
 19: /cgn2_6/_ptodata/2/pubpna/us60_PUBCOMB.seq:*

Database : Published Applications NA:
 1: /cgn2_6/_ptodata/2/pubpna/us07_PUBCOMB.seq:*
 2: /cgn2_6/_ptodata/2/pubpna/pct_new_pub.seq:*
 3: /cgn2_6/_ptodata/2/pubpna/us06_NEW_PUB.seq:*
 4: /cgn2_6/_ptodata/2/pubpna/us06_PUBCOMB.seq:*
 5: /cgn2_6/_ptodata/2/pubpna/us07_NEW_PUB.seq:*
 6: /cgn2_6/_ptodata/2/pubpna/us08_PUBCOMB.seq:*
 7: /cgn2_6/_ptodata/2/pubpna/us08_PUBCOMB.seq:*
 8: /cgn2_6/_ptodata/2/pubpna/us08_PUBCOMB.seq:*
 9: /cgn2_6/_ptodata/2/pubpna/us09B_PUBCOMB.seq:*
 10: /cgn2_6/_ptodata/2/pubpna/us09C_PUBCOMB.seq:*
 11: /cgn2_6/_ptodata/2/pubpna/us09_NEW_PUB.seq:*
 12: /cgn2_6/_ptodata/2/pubpna/us09_NEW_PUB.seq:*
 13: /cgn2_6/_ptodata/2/pubpna/us10A_PUBCOMB.seq:*
 14: /cgn2_6/_ptodata/2/pubpna/us10A_PUBCOMB.seq:*
 15: /cgn2_6/_ptodata/2/pubpna/us10B_PUBCOMB.seq:*
 16: /cgn2_6/_ptodata/2/pubpna/us10C_PUBCOMB.seq:*
 17: /cgn2_6/_ptodata/2/pubpna/us10_NEW_PUB.seq:*
 18: /cgn2_6/_ptodata/2/pubpna/us60_NEW_PUB.seq:*
 19: /cgn2_6/_ptodata/2/pubpna/us60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------|---------------------|
| 1 | 911 | 100.0 | 911 | 15 | US-10-069-434-4 |
| 2 | 759 | 83.3 | 1248 | 13 | US-10-220-120-4 |
| 3 | 745.8 | 81.9 | 1023 | 13 | US-10-43-802-19 |
| 4 | 745.8 | 81.9 | 3564 | 16 | US-10-108-260A-639 |
| 5 | 444.4 | 48.8 | 496 | 10 | US-09-918-995-14076 |
| 6 | 331.2 | 36.4 | 2785 | 9 | US-09-802-674-3 |
| 7 | 331.2 | 36.4 | 2785 | 15 | US-10-393-892-31 |
| 8 | 331.2 | 36.4 | 2785 | 15 | US-10-394-382-31 |
| 9 | 329.6 | 36.2 | 1233 | 15 | US-10-106-698-360 |
| 10 | 329.6 | 36.2 | 1244 | 15 | US-09-981-53-79 |
| 11 | 329.6 | 36.2 | 1244 | 15 | US-10-235-994-23 |
| 12 | 329.6 | 36.2 | 2826 | 15 | US-10-158-646-47 |
| 13 | 289.4 | 31.8 | 1459 | 9 | US-09-917-800A-1675 |
| 14 | 289.4 | 31.8 | 1459 | 16 | US-10-191-803-124 |

RESULT 1
US-10-069-434-4

; Sequence 4, Application US-10069434
; Publication No. US20030121061A1
; GENERAL INFORMATION:
; APPLICANT: INCYTE GENOMICS, INC.
; APPLICANT: THORNTON, Michael L.
; APPLICANT: RAMKUMAR, Jayalakshmi
; APPLICANT: TRIBOURLEY, Catherine M.
; APPLICANT: YUE, Henry
; APPLICANT: NGUYEN, Dannie L.
; APPLICANT: YAO, Monique G.
; APPLICANT: PATERSON, Chandra
; APPLICANT: GANDHI, Ameena R.
; APPLICANT: BUTFORD, Neil
; APPLICANT: THANGAVELU, Kavitha
; APPLICANT: BAUGHN, Mariah R.
; TITLE OF INVENTION: HUMAN LINES
; FILE REFERENCE: PI-0137 PCT
; CURRENT APPLICATION NUMBER: US-10/069_434
; CURRENT FILING DATE: 2002-07-20
; PRIOR APPLICATION NUMBER: 60/213,383 ; 60/215,544 ; 60/222,818
; NUMBER OF SEQ ID NOS: 6
; SEQ ID NO: 4
; LENGTH: 911
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: miBC_feature
; OTHER INFORMATION: Incyte ID No. US20030121061A1 63383333CB1
; US-10-069-434-4

ALIGNMENTS

Query Match 100.0%; Score 911; Length 911;
Best Local Similarity 100.0%; Pred. No. 1.3e-283;
Matches 911; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CGGAATTCCGCTCGAGTTCCACCCGGAGGACCATGTCAGGCTCAGCTGGGATAACGCC 60
 Db 1 CGGAAATTGGCTCGAGTTCCACCCGGAGGACCATGTCAGGCTCAGCTGGGATAACGCC 60
 QY 61 GAGGCAACCGTCTTAATCTGAAAGGAAATTTCCTTAATGGTGTATCGZRA 120
 Db 61 GAGGCAACCGTCTTAATCTGAAAGGAAATTTCCTTAATGGTGTATCGZRA 120
 QY 121 TCTCCAAATGGGATTTCCATTACTGAAAGGAAATTTCCTGACCACCTAAT 180
 Db 121 TCTCCAAATGGGATTTCCATTACTGAAAGGAAATTTCCTGACCACCTAAT 180
 QY 181 ATCAAGTATGACCCAAAGCTCAAGTAAATCATGAAACAGCGCCATTCCTCAAT 240
 Db 181 ATCAAGTATGACCCAAAGCTCAAGTAAATCATGAAACAGCGCCATTCCTCAAT 240
 QY 241 GACTTGATGACACAGAAACAANTCAGTTCTGGTGNGGTCTCTACTGGAAGCTAC 300
 Db 241 GACTTGATGACACAGAAACAANTCAGTTCTGGTGNGGTCTCTACTGGAAGCTAC 300
 QY 301 AGGTTAGCCAGGTTCACTTCACTGGGCTCAGTGGTACAGCGGCTCGAGCACATA 360
 Db 301 AGGTTAGCCAGGTTCACTTCACTGGGCTCAGTGGTACAGCGGCTCGAGCACATA 360
 QY 361 GTAGATGGACTGAGGTAGTGTGCGAGGTCATGTGTTCATGGAATTCAGAACATAC 420
 Db 361 GTATGGACTGAGGTAGTGTGCGAGGTCATGTGTTCATGGAATTCAGAACATAC 420
 QY 421 CCCAGCTTGTGAGGAGCTCATGAAACCAGATGGACCTGGCTCTGGAGTGTCTTA 480
 Db 421 CCCAGCTTGTGAGGAGCTCATGAAACCAGATGGACCTGGCTCTGGAGTGTCTTA 480
 QY 481 CAGATTGGCAACCTAACTCCAACTGCAAGGAAATTCTGACACTTGGATTCCATTAA 540
 Db 481 CAGATTGGCAACCTAACTCCAACTGCAAGGAAATTCTGACACTTGGATTCCATTAA 540
 QY 541 GAAAAGGTAAACAAACCTGAACTTCAAACTGCAAGGAAATTCTGACACTTGGATTCCATTAA 600
 Db 541 GAAAAGGTAAACAAACCTGAACTTCAAACTGCAAGGAAATTCTGACACTTGGATTCCATTAA 600
 QY 601 TGGGACTACTGGACATATCCGGTCTTACACTTCAACCTCTTGAGAGTGTCTCA 660
 Db 601 TGGGACTACTGGACATATCCGGTCTTACACTTCAACCTCTTGAGAGTGTCTCA 660
 QY 661 TGGGATTGTTAAAGCAACCTAAACATGAGCTCTCAACAGTGGCCAATTTCGCCT 720
 Db 661 TGGGATTGTTAAAGCAACCTAAACATGAGCTCTCAACAGTGGCCAATTTCGCCT 720
 QY 721 CTCTGTGCAAGCGGAGGTGANGCAGCTTCTGTGATAAGTGTCACTCTGTC 780
 Db 721 CTCTGTGCAAGCGGAGGTGANGCAGCTTCTGTGATAAGTGTCACTCTGTC 780
 QY 781 CCCAGGCTGGGGCACTGGTACATCTGGCTTAATGGCCTCCAACTCCTGGACTCA 840
 Db 781 CCCAGGCTGGGGCACTGGTACATCTGGCTTAATGGCCTCCAACTCCTGGACTCA 840
 QY 901 CAGGGGGAAA 911
 Db 901 CAGGGGGAAA 911

; APPLICANT: SPIRO, Peter A.
; APPLICANT: BANVILLE, Steven C.
; APPLICANT: ;
; APPLICANT: SHAH, Purvi
; APPLICANT: CHALUP, Michael S.
; APPLICANT: CHENG, Simon C.
; APPLICANT: CHEN, Alice
; APPLICANT: D'SA, Steven A.
; APPLICANT: AMBHEY, Stefan
; APPLICANT: DAHL, Christopher R.
; APPLICANT: DAM, Tam C.
; APPLICANT: DANIELS, Susan E.
; APPLICANT: DUFOUR, Gerard E.
; APPLICANT: FLORES, Vincent
; APPLICANT: FONG, Willy T.
; APPLICANT: GREENWALT, Lila B.
; APPLICANT: HILLMAN, Jennifer L.
; APPLICANT: JONES, Anissa L.
; APPLICANT: LIU, Tommy F.M.
; APPLICANT: ROBBERY, Ann M.
; APPLICANT: ROSEN, Bruce H.
; APPLICANT: RUSSO, Frank D.
; APPLICANT: STOCKDRISHER, Theresa K.
; APPLICANT: DAFO, Abel
; APPLICANT: WRIGHT, Rachel J.
; APPLICANT: YAH, Pierre E.
; APPLICANT: YU, Jimmy Y.
; APPLICANT: BRADLEY, Diana L.
; APPLICANT: BRATCHER, Shawn R.
; APPLICANT: CHEN, Wensheng
; APPLICANT: COHEN, Howard J.
; APPLICANT: HOLSON, David M.
; APPLICANT: LINCOLN, Stephen E.
; APPLICANT: JACKSON, Stuart
; TITLE OF INVENTION: MOLECULES FOR DIAGNOSTICS AND THERAPEUTICS
; CURRENT APPLICATION NUMBER: US/10/20,120
; CURRENT FILING DATE: 2002-08-26
; PRIOR APPLICATION NUMBER: 60/184,777; 60/184,797; 60/184,698; 60/184,770; 60/184,774,
; FILE REFERENCE: PT-1113 PCT
; SEQ ID NO: 4
; LENGTH: 1248
; ORGANISM: Homo sapiens
; FEATURE: misc feature
; NAME/KEY: unsure
; LOCATION: 1148, 1234
; OTHER INFORMATION: a, t, c, g, or other
; US-10-220-120-4

Query Match 83-3%; Score 759; DB 13; Length 1248;
 Best Local Similarity 10.0%; Pred No. 2.5e-234;
 Matches 759; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CGGAAATTGGCTGGAGTTCACCCGGAGGACATGTCAGGAGCTCACTGGGATACCG 60
 Db 15 CGGAAATTGGCTGGAGTTCACCCGGAGGACATGTCAGGAGCTCACTGGGATACCG 74

RESULT 2
 US-10-220-120-4
 Sequence 4, Application US/10220120
 ; Publication No. US2004004253A1
 ; GENERAL INFORMATION:
 ; APPLICANT: INCYTE GENOMICS, INC.
 ; APPLICANT: PANZER, Scott R.

61 GAGCACAAACGGCTCATTCACTGGAGGAATTTCGCCATTGCTGATGGTGTAGCAA 120
 75 GAGCACAAACGGCTCATTCACTGGAGGAATTTCGCCATTGCTGATGGTGTAGCAA 134
 Qy 121 TCTCCAAATTGAGATAAAGAACAAAGTAGAAATATGACTTTCCCTCCGACCCTTGT 180
 Db 135 TCTCCAAATTGAGATAAAGAACAAAGTAGAAATATGACTTTCCCTCCGACCCTTGT 194
 Qy 181 ATCAAGTATGACCCAAAGCTAGCTAACATGAAACCCGGCATTCCTTCATGT 240
 Db 195 ATCAAGTATGACCCAAAGCTAGCTAACATGAAACCCGGCATTCCTTCATGT 254
 Qy 241 GACTTGTAGACACAGAAACAAATGATTTGCGCTGGTGTCTCACTGGAGCTAC 300
 Db 255 GACTTGTAGACACAGAAACAAATGATTTGCGCTGGTGTCTCACTGGAGCTAC 314
 Qy 301 AGTTTAAGGAGGTTAACCTTCACTGGGGTCCGCTGATGACCAAGGCTCGAGACATA 360
 Db 315 AGTTTAAGGAGGTTAACCTTCACTGGGGTCCGCTGATGACCAAGGCTCGAGACATA 374
 Qy 361 GTAGATGGAGTGAGGTATGGCTCAGACCTTCACTGGGGTCCGCTGATGACAAATAC 420
 Db 375 GTAGATGGAGTGAGGTATGGCTCAGACCTTCACTGGGGTCCGCTGATGACAAATAC 434
 Qy 421 CCCAGTTTGTGGAGGCTCATGACCAAGATGGACTGGCTGTCCTGGAGTTTTA 480
 Db 435 CCCAGTTTGTGGAGGCTCATGACCAAGATGGACTGGCTGTCCTGGAGTTTTA 494
 Qy 481 CAGATGGTGAACCTTAATTCCAACCTGAAAGATAACTGACACTTGGATTCCATTAA 540
 Db 495 CAGATGGTGAACCTTAATTCCAACCTGAAAGATAACTGACACTTGGATTCCATTAA 554
 Qy 541 GAAAAGGTTAACAAACTCGATTCAAAATTGACCTTGTCTCCTGCACCATTC 600
 Db 555 GAAAAGGTTAACAAACTCGATTCAAAATTGACCTTGTCTCCTGCACCATTC 614
 Qy 601 TGGGACTACTGGACATATCCTCGTTCTCTACAGTTCACTGGAGGTGTCA 660
 Db 615 TGGGACTACTGGACATATCCTCGTTCTCTACAGTTCACTGGAGGTGTCA 674
 Qy 661 TGATTTAAAGCAACTTAACATCAGCTCTAACAGCTGGCAAAATTTCGCA 720
 Db 675 TGATTTAAAGCAACTTAACATCAGCTCTAACAGCTGGCAAAATTTCGCA 734
 Qy 721 CTCTGTGACACGGGAGGGTGAAGCACAGCTTTCCTG 759
 Db 735 CTCTGTGACACGGGAGGGTGAAGCACAGCTTTCCTG 773

RESULT 3
 US-10-433-802-19
 ; Sequence 19 , Application US/10433802
 ; Publication No. US20040063115A1
 ; GENERAL INFORMATION:
 ; APPLICANT: TANG, Y. Tom; GRIFFIN, Jennifer A.;
 ; APPLICANT: YUE, Henry; LEE, Ernestine A.;
 ; APPLICANT: BAUGHN, Mariah R.; DUGGAN, Brendan M.;
 ; APPLICANT: CHAWLA, Narinder K.; LEE, Sally;
 ; APPLICANT: RAMKOMAR, Chaitanya; WARREN, Bridget A.;
 ; APPLICANT: GANDHI, Ameena R.; LU, Dyoung Aina M.;
 ; APPLICANT: LIU, Yan; YAO, Monique G.;
 ; APPLICANT: DING, Li; TRIBOUTEY, Catherine M.;
 ; APPLICANT: SANJANWALA, Madhusudan M.;
 ; APPLICANT: ARVIZU, Chandra S.; JACKSON, Jennifer L.
 ; TITLE OF INVENTION: ENZYME
 ; FILE REFERENCE: PI-0316 PCT
 ; CURRENT APPLICATION NUMBER: US/10/433, 802
 ; CURRENT FILING DATE: 2003-06-04
 ; PRIOR APPLICATION NUMBER: PCT/US01/47432
 ; PRIOR FILING DATE: 2001-12-04
 ; PRIOR APPLICATION NUMBER: US 60/251, 824
 ; PRIOR FILING DATE: 2000-12-07

; PRIOR APPLICATION NUMBER: US 60/254, 312
 ; PRIOR FILING DATE: 2000-12-08;
 ; PRIOR APPLICATION NUMBER: US 60/255, 773
 ; PRIOR FILING DATE: 2000-12-14
 ; PRIOR APPLICATION NUMBER: US 60/256, 188
 ; PRIOR FILING DATE: 2000-12-15
 ; PRIOR APPLICATION NUMBER: US 60/255, 940
 ; PRIOR FILING DATE: 2000-12-15
 ; PRIOR APPLICATION NUMBER: US 60/257, 498
 ; PRIOR FILING DATE: 2000-12-21
 ; PRIOR APPLICATION NUMBER: US 60/262, 839
 ; PRIOR FILING DATE: 2001-01-19
 ; PRIOR APPLICATION NUMBER: US 60/264, 402
 ; PRIOR FILING DATE: 2001-01-26
 ; NUMBER OF SEQ ID NOS: 36
 ; SOFTWARE: PBRLL Program
 SEQ ID NO 19
 LENGTH: 1023
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; NAME/KEY: misc_feature
 ; OTHER INFORMATION: Incyte ID No. US20040063115A1 8159895CB1
 ; OTHER INFORMATION: US-10-433-802-19

Query Match Score 745 : 8; DB 13; Length 1023;
 Best Local Similarity 97.2%; Pred. No. 4,1e-230;
 Matches 759; Conservative 0; Mismatches 22; Indels 0; Gaps 0;

Qy 17 TTCCACCCGGGGACCATGTCAGGCTCAGCTGGGATACCGCGAGCACACGGTCTCA 76
 Db 1 TTCCACCCGGGGACCATGTCAGGCTCAGCTGGGATACCGCGAGCACACGGTCTCA 60

Qy 77 TTACATGGAGGAATTTCCTATTGCTGATGGTATCAGCAATTCCTGAGATTGAGATA 136
 Db 61 TTACATGGAGGAATTTCCTATTGCTGATGGTATCAGCAATTCCTGAGATTGAGATA 120

Qy 137 AAACCAAAGAACTGAATAATGACTCTTCCTCGAACCTTAGTATCAAGTATGACCAA 196
 Db 121 AAACCAAAGAACTGAATAATGACTCTTCCTCGAACCTTAGTATCAAGTATGACCAA 180

Qy 197 GCTCAGCTAAATCATAGCAACAGGGGCCATTCTTCATGACTTTGACTTTGATGACACAG 256
 Db 181 GCTCAGCTAAATCATAGCAACAGGGGCCATTCTTCATGACTTTGATGACACAG 240

Qy 257 AGACAAATCATGCTTCCTGCTGATGACCTTGTGACCTTGTGACCTTGTGACCTTGTGACCT 316
 Db 241 AGACAAATCATGCTTCCTGCTGATGACCTTGTGACCTTGTGACCTTGTGACCT 300

Qy 317 ACCTTCACTGGGGTCCGCTGATGACCAAGGGTCCGAGCACTAGTATGAGTGA 376
 Db 301 ACCTTCACTGGGGTCCGCTGATGACCAAGGGTCCGAGCACTAGTATGAGTGA 360

Qy 377 ATGCTGAGAGCTCCATGTTGTTCACTGAAATTGAGCAATCCCGAGCTTGTGAGG 436
 Db 361 ATGCTGAGAGCTCCATGTTGTTCACTGAAATTGAGCAATCCCGAGCTTGTGAGG 420

Qy 437 CAGCTCATGAAACGAGGACTGGCTCTCTGGAGTGTGTTTACAGATTGTGAGCTA 496
 Db 421 CAGCTCATGAAACGAGGACTGGCTCTCTGGAGTGTGTTTACAGATTGTGAGCTA 480

Qy 497 ATTCCCAACTGCAAAAGATACTGACACTTGGATTCATTAAGAAAGGTTAAAGAAA 556
 Db 481 ATTCCCAACTGCAAAATTGACCTTGTGACCTTGGGACTACTGGACAT 540

Qy 557 CTCGATTCACAAATTGACCTTGTGCTCTCCACCATCTGGGACTACTGGACAT 616
 Db 541 CTCGATTCACAAATTGACCTTGTGCTCTCCACCATCTGGGACTACTGGACAT 600

Qy 617 ATCCTGGTCTCTTAAGCTCCACCTCTGAGGTGTGACATGAGTTGTTAAGGC 676
 Db 601 ATCCTGGTCTCTTAAGCTCCACCTCTGAGGTGTGACATGAGTTGTTAAGGC 660

| | | | |
|---|---|----|--|
| Qy | 677 AACCTATAAACATCAGCTCTCACAGCTGGCCAATTTCGCAAGTCCCTGTGCACAGGG 736 | Db | 575 CTGGATTACAATTTGACCTATTTGCTCTGGTCCACCATTCCTGGACTAATGGACAT 634 |
| Db | 651 AACCTATAAACATCAGCTCTCACAGCTGGCCAATTTCGCAAGGG 720 | Qy | 617 ATCGTGGTCTCTCACATTCCACCTCTCTGAGATGCTACATGAAATTGTTTAAGC 676 |
| Qy | 737 AGGCTGAACGAGCTTCTGTGATAGTCACCTGTGACCTGGCTGGGCCA 796 | Db | 635 ATCGTGGTCTCTCACATTCCACCTCTGAGATGCTACATGAAATTGTTTAAGC 694 |
| Db | 721 AGGCTGAACGAGCTTCTGTGAGCAATTCCACGGCCACACGGCTTAAGGCC 780 | Qy | 677 AACCTATAAACATCAGCTCTCACAGCTGGCCAATTTCGCAAGGG 736 |
| Qy | 797 G 797 | Db | 695 AACCTATAAACATCAGCTCTCACAGCTGGCCAATTTCGCAAGGG 754 |
| Db | 781 G 781 | Qy | 737 AGGCTGAAGCAGCTTCCTGTGATAGACTCTACCTGTCACCAGTGCAATTTCGAGTCTCTGTGACAGGGCA 796 |
| RESULT 4 | | | |
| US-10-108-260A-639 | | | |
| ; sequence 639, Application US/10108260A | | | |
| ; Publication No. US20040005560A1 | | | |
| ; GENERAL INFORMATION: | | | |
| ; APPLICANT: HELIX RESEARCH INSTITUTE | | | |
| ; TITLE OF INVENTION: NO. US20040005560A1 full length cDNA | | | |
| ; FILE REFERENCE: H_A0106 | | | |
| ; CURRENT APPLICATION NUMBER: US/10/108,260A | | | |
| ; CURRENT FILING DATE: 2002-03-27 | | | |
| ; NUMBER OF SEQ ID NOS: 5458 | | | |
| ; SOFTWARE: Patentin Ver. 2.1 | | | |
| ; SEQ ID NO: 639 | | | |
| ; LENGTH: 3564 | | | |
| ; TYPE: DNA | | | |
| ; ORGANISM: Homo sapiens | | | |
| US-10-108-260A-639 | | | |
| Query Match Score 81.9%; Best Local Similarity 97.2%; Matches 759; Conservative 0; Mismatches 22; Indels 0; Gaps 0; | | | |
| Qy | 17 TTCCACCCGAGGGACCATGTCGAGGCTAGCTGGGTATACCGCGAGACAAAGGTCTTA 76 | Qy | 17 TTCACTGAGGATTTCCTATTGTGATGCTGTGATCTCAATGGAGATA 136 |
| Db | 35 TTCCACCCGAGGGACCATGTCGAGGCTAGCTGGGTATACCGCGAGACAAAGGTCTTA 94 | Qy | 95 TTCACTGAGGATTTCCTATTGTGATGCTGTGATCTCAATGGAGATA 154 |
| Qy | 137 AAACCAAAGTAAATAATGACTCTTCCGACCACTTAGTCAAGTAGTACCCAA 196 | Db | 155 AACCAAAGTAAATAATGACTCTTCCGACCACTTAGTCAAGTAGTACCCAA 214 |
| Db | 197 GCTCAGCTAAATCATCAGAACAGCGCCAATTCTCAATTGTGACTTTGATGACAG 256 | Qy | 197 GCTCAGCTAAATCATCAGAACAGCGCCAATTCTCAATTGTGACTTTGATGACAG 256 |
| Db | 215 GCTCAGCTAAATCATCAGAACAGCGCCAATTCTCAATTGTGACTTTGATGACAG 274 | Qy | 257 AGAACAAATCAGTCTGCTGGTCTCTCACTGGAGCTACGGCAGGTTC 316 |
| Qy | 275 AGAACAAATCAGTCTGCTGGTCTCTCACTGGAGCTACGGCAGGTTC 334 | Db | 275 AGAACAAATCAGTCTGCTGGTCTCTCACTGGAGCTACGGCAGGTTC 334 |
| Qy | 317 ACCTTCATCTGGGTCCGTGATACCCGGCTCGAGCAATTAGTAGATGGATGAGCT 376 | Qy | 317 ACCTTCATCTGGGTCCGTGATACCCGGCTCGAGCAATTAGTAGATGGATGAGCT 376 |
| Db | 335 ACCTTCACTGGGTCCGTGATACCCGGCTCGAGCAATTAGTAGATGGATGAGCT 394 | Db | 335 ACCTTCACTGGGTCCGTGATACCCGGCTCGAGCAATTAGTAGATGGATGAGCT 394 |
| Qy | 377 ATGCTGCAAGACTCATGTTGTCACGGAAATTCAAGAAATTCCCGCTTGTGG 436 | Qy | 437 CAGCTCATGAAACGATGACTGCTGTTGGAGCTTTCAGATTGCTGAACTTA 496 |
| Db | 395 ATGCTGCAAGACTCATGTTGTCACGGAAATTCAAGAAATTCCCGCTTGTGG 454 | Db | 455 CAGCTCATGAAACGATGACTGCTGTTGGAGCTTTCAGATTGCTGAACTTA 514 |
| Qy | 497 ATGCCCAACTGCAAAGATACTGACACTTGTGACTTTGATGAGGTAAACAA 556 | Qy | 497 ATGCCCAACTGCAAAGATACTGACACTTGTGACTTTGATGAGGTAAACAA 556 |
| Db | 515 ATTCCCCAACTGCAAAGATACTGACACTTGTGACTTTGATGAGGTAAACAA 574 | Db | 515 ATTCCCCAACTGCAAAGATACTGACACTTGTGACTTTGATGAGGTAAACAA 574 |
| Qy | 557 CTGGATTACAATTTGACCTATTGTCTGTCACCGCTTGTGAGCT 616 | Qy | 557 CTGGATTACAATTTGACCTATTGTCTGTCACCGCTTGTGAGCT 616 |

RESULT 6
US-09-802-674-3
; Sequence 3, Application US/09802674
; Patent No. US20042088A1
; GENERAL INFORMATION:
; APPLICANT: Macina, Roberto A.
; ATTORNEY: Pinerit, Alejandra
; APPLICANT: Sun, Yongming
; TITLE OF INVENTION: Method of Diagnosing, Monitoring, Staging, Imaging and Treating Gastrointestinal Cancer
; FILE REFERENCE: DEX-0142
; CURRENT APPLICATION NUMBER: US/09/802,674
; CURRENT FILING DATE: 2001-03-09
; PRIOR APPLICATION NUMBER: 60/188,061
; PRIOR FILING DATE: 2000-03-09
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
; LENGTH: 2785
; TYPE: DNA: Homo sapiens
; ORGANISM: Homo sapiens
; US-09-802-674-3

Query Match Score 331.2; DB 9; Length 2785;
Best Local Similarity 64.5%; Pred. No. 1.2e-95; Indels 0; Gaps 0;
Matches 495; Conservative 273; Indels 0; Gaps 0;
US-10-393-892-31

RESULT 7
US-10-393-892-31
; Sequence 31, Application US/10393892
; Publication No. US20030186302A1
; GENERAL INFORMATION:
; APPLICANT: WANG, YIXIN
; TITLE OF INVENTION: COLORECTAL CANCER DIAGNOSTICS
; FILE REFERENCE: CDS 267 US NP
; CURRENT APPLICATION NUMBER: US/10/393,892
; CURRENT FILING DATE: 2003-03-21
; PRIOR APPLICATION NUMBER: 60/368,798
; PRIOR FILING DATE: 2002-03-29
; NUMBER OF SEQ ID NOS: 49
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 31
; LENGTH: 2785
; TYPE: DNA
; ORGANISM: human
; US-10-393-892-31

Query Match Score 331.2; DB 15; Length 2785;
Best Local Similarity 64.5%; Pred. No. 1.2e-55; Indels 0; Gaps 0;
Matches 495; Conservative 273; Indels 0; Gaps 0;

Query Match Score 36.4%; Best Local Similarity 64.5%; Pred. No. 1.2e-55; Indels 0; Gaps 0;
Matches 495; Conservative 0; Mismatches 273; Indels 0; Gaps 0;

Query Match Score 36.4%; Best Local Similarity 64.5%; Pred. No. 1.2e-55; Indels 0; Gaps 0;
Matches 495; Conservative 0; Mismatches 273; Indels 0; Gaps 0;

Query Match Score 30 GACCAGTCAGGCTCACTGGGGATACCGGAGCACACGGTCCATTACTGGAGGA 89
Best Local Similarity 64.5%; Pred. No. 1.2e-55; Indels 0; Gaps 0;
Matches 495; Conservative 0; Mismatches 273; Indels 0; Gaps 0;

Query Match Score 30 GACCAGTCAGGCTCACTGGGGATACCGGAGCACACGGTCCATTACTGGAGGA 89
Best Local Similarity 64.5%; Pred. No. 1.2e-55; Indels 0; Gaps 0;
Matches 495; Conservative 0; Mismatches 273; Indels 0; Gaps 0;

Query Match Score 1044 GATAATGCCAAGTCCAGACTGGGATATGTGACA 1103
Best Local Similarity 64.5%; Pred. No. 1.2e-55; Indels 0; Gaps 0;
Matches 495; Conservative 0; Mismatches 273; Indels 0; Gaps 0;

Query Match Score 1044 GATAATGCCAAGTCCAGACTGGGATATGTGACA 1103
Best Local Similarity 64.5%; Pred. No. 1.2e-55; Indels 0; Gaps 0;
Matches 495; Conservative 0; Mismatches 273; Indels 0; Gaps 0;

Query Match Score 90 ATTTTCCTATTGCTGATGGGATATGCAATCTCCATGAGATAAAAGAGT 149
Best Local Similarity 64.5%; Pred. No. 1.2e-55; Indels 0; Gaps 0;
Matches 495; Conservative 0; Mismatches 273; Indels 0; Gaps 0;

Query Match Score 1104 GCTGTATCCCATGCCATGGAAATAAACCATTCCCTGTGTATTAACCAAGTGAAC 1163
Best Local Similarity 64.5%; Pred. No. 1.2e-55; Indels 0; Gaps 0;
Matches 495; Conservative 0; Mismatches 273; Indels 0; Gaps 0;

Query Match Score 150 GAAATATGACTCTTCCCAGAACCTTAAGTCAAGTATGACCCAAAGCTCAGCTAAAT 209
Best Local Similarity 64.5%; Pred. No. 1.2e-55; Indels 0; Gaps 0;
Matches 495; Conservative 0; Mismatches 273; Indels 0; Gaps 0;

Query Match Score 1164 CAAACATGACCCCTGAAACCTTATAGTGTCTAACCCAGCCAAG 1223
Best Local Similarity 64.5%; Pred. No. 1.2e-55; Indels 0; Gaps 0;
Matches 495; Conservative 0; Mismatches 273; Indels 0; Gaps 0;

Query Match Score 210 CATCAGGAAAGCGGCCATTCCCTCAATGTTGACTTTGATGACACAGAGAACATTGCT 269
Best Local Similarity 64.5%; Pred. No. 1.2e-55; Indels 0; Gaps 0;
Matches 495; Conservative 0; Mismatches 273; Indels 0; Gaps 0;

Query Match Score 1224 AATTATGATGTGGGATATTCCATGTAATTGAGGAACCATAAACGATCT 1283
Best Local Similarity 64.5%; Pred. No. 1.2e-55; Indels 0; Gaps 0;
Matches 495; Conservative 0; Mismatches 273; Indels 0; Gaps 0;

Query Match Score 270 TCTGCGTGTGGCTCTCACTGGAACTCAGGTACGGTTACCTCTACGGG 329
Best Local Similarity 64.5%; Pred. No. 1.2e-55; Indels 0; Gaps 0;
Matches 495; Conservative 0; Mismatches 273; Indels 0; Gaps 0;

Query Match Score 1284 GCTGAAAGGTGTCCTTCATGTTGAGCTATGGCTCTTCAAGGCTAAGCGATCTGGG 1343
Best Local Similarity 64.5%; Pred. No. 1.2e-55; Indels 0; Gaps 0;
Matches 495; Conservative 0; Mismatches 273; Indels 0; Gaps 0;

Query Match Score 330 GTCGGCTGATGACCAAGCTCCGAGCACATAGTGAAGTATGCTGAGCT 389
Best Local Similarity 64.5%; Pred. No. 1.2e-55; Indels 0; Gaps 0;
Matches 495; Conservative 0; Mismatches 273; Indels 0; Gaps 0;

Query Match Score 1344 CAGTACAATGAGCATGTTGAGAACATACAGTGGATGAGCTAAATTCGCGAGCT 1403
Best Local Similarity 64.5%; Pred. No. 1.2e-55; Indels 0; Gaps 0;
Matches 495; Conservative 0; Mismatches 273; Indels 0; Gaps 0;

Query Match Score 390 CCATGTTGACTGAGGATTCAGATGGGACTCTATCCCACTTGTGAGGCTCATGAA 509
Best Local Similarity 64.5%; Pred. No. 1.2e-55; Indels 0; Gaps 0;
Matches 495; Conservative 0; Mismatches 273; Indels 0; Gaps 0;

Query Match Score 1404 TCAGTGTGCTACTGGAAATTCTGAAACTACTCCAGCTTGTGAAAGTCCAAAGGC 1463
Best Local Similarity 64.5%; Pred. No. 1.2e-55; Indels 0; Gaps 0;
Matches 495; Conservative 0; Mismatches 273; Indels 0; Gaps 0;

Query Match Score 450 AGATGGACTGGCTGCTGGAGTTACAGATGGTGAACCTTATCCAACTGTCGA 509
Best Local Similarity 64.5%; Pred. No. 1.2e-55; Indels 0; Gaps 0;
Matches 495; Conservative 0; Mismatches 273; Indels 0; Gaps 0;

Query Match Score 1464 TGAATGACTGAGCTGCTTGTGAGGTGGCTTGTGAAAGTCTGCAAA 1523
Best Local Similarity 64.5%; Pred. No. 1.2e-55; Indels 0; Gaps 0;
Matches 495; Conservative 0; Mismatches 273; Indels 0; Gaps 0;

Query Match Score 510 AAAGATTACTGACACTTGATTCATTAAGAAAAGCTTCAAAACTGATTCAA 569
Best Local Similarity 64.5%; Pred. No. 1.2e-55; Indels 0; Gaps 0;
Matches 495; Conservative 0; Mismatches 273; Indels 0; Gaps 0;

Query Match Score 570 TTTGACCTATTGCTGCTGCTTGGACTACTGGACATATCCNGGTTCT 629
Best Local Similarity 64.5%; Pred. No. 1.2e-55; Indels 0; Gaps 0;
Matches 495; Conservative 0; Mismatches 273; Indels 0; Gaps 0;

RESULT 8
US-10-394-382-31
; Sequence 31, Application US/10394382
; PUBLICATION NO. US20030186303A1

; APPLICANT: Wang, Yixin
; TITLE OF INVENTION: COLORECTAL CANCER DIAGNOSTICS
; FILE REFERENCE: CDS 266 US NP
; CURRENT APPLICATION NUMBER: US/10/394,382
; CURRENT FILING DATE: 2003-03-21
; PRIOR APPLICATION NUMBER: 60/368,687
; PRIOR FILING DATE: 2002-03-29
; NUMBER OF SEQ ID NOS: 49
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 31
; LENGTH: 2785
; TYPE: DNA
; ORGANISM: human
; US-10-394-382-31

Query Match Score 331.2; DB 15; Length 2785;
Best Local Similarity 64.5%; Pred. No. 1.2e-95;
Matches 495; Conservative 0; Mismatches 273; Indels 0; Gaps 0;

Qy 30 GACCATGTCAGGGTCAAGCTGGGATAACCGGAGCACACGGTCTATTACTGGAAAGCA 89
Db 1044 GATAATGGGAAGTCAGACTGGGATAATGACAATAATGGGAAATGGGAA 1103
; LOCATION: (1122) .. (1122)
; OTHER INFORMATION: n equals a,t,g, or c
; US-10-394-382-31

Query Match Score 329.6; DB 15; Length 1233;
Best Local Similarity 64.3%; Pred. No. 2.4e-95;
Matches 494; Conservative 27; Mismatches 0; Indels 0; Gaps 0;

Qy 30 GACCATGTCAGGGTCAAGCTGGGATAACCGGAGCACACGGTCTATTACTGGAAAGCA 89
Db 73 GATAATGGGAAGTCAGACTGGGATAATGACAATAATGGGAAATGGGAA 132
; LOCATION: (1122) .. (1122)
; OTHER INFORMATION: n equals a,t,g, or c
; US-10-394-382-31

Query Match Score 329.6; DB 15; Length 1233;
Best Local Similarity 64.3%; Pred. No. 2.4e-95;
Matches 494; Conservative 27; Mismatches 0; Indels 0; Gaps 0;

Qy 30 GACCATGTCAGGGCATTCTGCAGACTGGGATAACCGGAGCACACGGTCTATTACTGGAAAGCA 89
Db 1164 CAACATGACACCTCTGAAACCTATAGTGTTCCACAAACGGCAAAAGCA 1223
; LOCATION: (1122) .. (1122)
; OTHER INFORMATION: n equals a,t,g, or c
; US-10-394-382-31

Query Match Score 329.6; DB 15; Length 1233;
Best Local Similarity 64.3%; Pred. No. 2.4e-95;
Matches 494; Conservative 27; Mismatches 0; Indels 0; Gaps 0;

Qy 30 GACCATGTCAGGGCATTCTGCAGACTGGGATAACCGGAGCACACGGTCTATTACTGGAAAGCA 89
Db 1224 AATPATCATGTCAGGGCATTCTGCAGACTGGGATAACCGGAGCACACGGTCTATTACTGGAAAGCA 1283
; LOCATION: (1122) .. (1122)
; OTHER INFORMATION: n equals a,t,g, or c
; US-10-394-382-31

Query Match Score 329.6; DB 15; Length 1233;
Best Local Similarity 64.3%; Pred. No. 2.4e-95;
Matches 494; Conservative 27; Mismatches 0; Indels 0; Gaps 0;

Qy 30 GACCATGTCAGGGCATTCTGCAGACTGGGATAACCGGAGCACACGGTCTATTACTGGAAAGCA 89
Db 1284 GCTGAAAGGTGGTCTGCAATGGAAATAACGTTCAATGGATAAACCGTGAAC 1343
; LOCATION: (1122) .. (1122)
; OTHER INFORMATION: n equals a,t,g, or c
; US-10-394-382-31

Query Match Score 329.6; DB 15; Length 1233;
Best Local Similarity 64.3%; Pred. No. 2.4e-95;
Matches 494; Conservative 27; Mismatches 0; Indels 0; Gaps 0;

Qy 30 GACCATGTCAGGGCATTCTGCAGACTGGGATAACCGGAGCACACGGTCTATTACTGGAAAGCA 89
Db 1344 CAGTACAATGGACATGGTCACTGGGATAACCGGAGCACACGGTCTATTACTGGAAAGCA 1403
; LOCATION: (1122) .. (1122)
; OTHER INFORMATION: n equals a,t,g, or c
; US-10-394-382-31

Query Match Score 329.6; DB 15; Length 1233;
Best Local Similarity 64.3%; Pred. No. 2.4e-95;
Matches 494; Conservative 27; Mismatches 0; Indels 0; Gaps 0;

Qy 30 GACCATGTCAGGGCATTCTGCAGACTGGGATAACCGGAGCACACGGTCTATTACTGGAAAGCA 89
Db 1404 TCAGCTGTAAGTCAGGGCATTCTGCAGACTGGGATAACCGGAGCACACGGTCTATTACTGGAAAGCA 1463
; LOCATION: (1122) .. (1122)
; OTHER INFORMATION: n equals a,t,g, or c
; US-10-394-382-31

Query Match Score 329.6; DB 15; Length 1233;
Best Local Similarity 64.3%; Pred. No. 2.4e-95;
Matches 494; Conservative 27; Mismatches 0; Indels 0; Gaps 0;

Qy 30 GACCATGTCAGGGCATTCTGCAGACTGGGATAACCGGAGCACACGGTCTATTACTGGAAAGCA 89
Db 1450 AGATGGCACTGGCTGCTGGGAGTTACAGGTTACAGGTTACAGGTTACAGGTTACAGGTTACAGGTCAGGG 329
; LOCATION: (1122) .. (1122)
; OTHER INFORMATION: n equals a,t,g, or c
; US-10-394-382-31

RESULT 9
US-10-106-698-360
; Sequence 360, Application US/10106698
; Publication No. US2003010969A1
; GENERAL INFORMATION:
; APPLICANT: Ruben et al.
; TITLE OF INVENTION: Colon and Colon Cancer Associated Polynucleotides and Polypeptides
; FILE REFERENCE: PA005P1
; CURRENT APPLICATION NUMBER: US/10/106,698
; CURRENT FILING DATE: 2002-03-27
; PRIORITY APPLICATION NUMBER: PCT/US00/26524
; PRIORITY FILING DATE: 2000-09-28
; PRIORITY APPLICATION NUMBER: US 60/157,137
; PRIORITY FILING DATE: 1999-09-29
; PRIORITY APPLICATION NUMBER: US 60/163,280
; PRIORITY FILING DATE: 1999-11-03
; NUMBER OF SEQ ID NOS: 8564
; SOFTWARE: PatentIn Ver. 3.0
; SEQ ID NO 360
; LENGTH: 1233
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE: miss feature
; NAME/KEY: miss feature
; LOCATION: (1122) .. (1122)
; OTHER INFORMATION: n equals a,t,g, or c
; US-10-106-698-360

Query Match Score 329.6; DB 15; Length 1233;
Best Local Similarity 64.3%; Pred. No. 2.4e-95;
Matches 494; Conservative 27; Mismatches 0; Indels 0; Gaps 0;

Qy 30 GACCATGTCAGGGCATTCTGCAGACTGGGATAACCGGAGCACACGGTCTATTACTGGAAAGCA 89
Db 73 GATAATGGGAAGTCAGACTGGGATAATGACAATAATGGGAAATGGGAA 132
; LOCATION: (1122) .. (1122)
; OTHER INFORMATION: n equals a,t,g, or c
; US-10-106-698-360

Query Match Score 329.6; DB 15; Length 1233;
Best Local Similarity 64.3%; Pred. No. 2.4e-95;
Matches 494; Conservative 27; Mismatches 0; Indels 0; Gaps 0;

Qy 30 GACCATGTCAGGGCATTCTGCAGACTGGGATAACCGGAGCACACGGTCTATTACTGGAAAGCA 89
Db 90 ATTTCCTCTATTGCTGATGGGATACTCCAAATTGAGATAAAACCAAGAGGT 149
; LOCATION: (1122) .. (1122)
; OTHER INFORMATION: n equals a,t,g, or c
; US-10-106-698-360

Query Match Score 329.6; DB 15; Length 1233;
Best Local Similarity 64.3%; Pred. No. 2.4e-95;
Matches 494; Conservative 27; Mismatches 0; Indels 0; Gaps 0;

Qy 30 GACCATGTCAGGGCATTCTGCAGACTGGGATAACCGGAGCACACGGTCTATTACTGGAAAGCA 89
Db 133 GCTGATCCCATGCCAACTGGAAATAACGTTCAATGGATAAAACCGTGAAC 192
; LOCATION: (1122) .. (1122)
; OTHER INFORMATION: n equals a,t,g, or c
; US-10-106-698-360

Query Match Score 329.6; DB 15; Length 1233;
Best Local Similarity 64.3%; Pred. No. 2.4e-95;
Matches 494; Conservative 27; Mismatches 0; Indels 0; Gaps 0;

Qy 30 GACCATGTCAGGGCATTCTGCAGACTGGGATAACCGGAGCACACGGTCTATTACTGGAAAGCA 89
Db 150 GAATATGACTCTCCCTGCAGACTGGGATAACCGGAGCACACGGTCTATTACTGGAAAGCA 209
; LOCATION: (1122) .. (1122)
; OTHER INFORMATION: n equals a,t,g, or c
; US-10-106-698-360

Query Match Score 329.6; DB 15; Length 1233;
Best Local Similarity 64.3%; Pred. No. 2.4e-95;
Matches 494; Conservative 27; Mismatches 0; Indels 0; Gaps 0;

Qy 30 GACCATGTCAGGGCATTCTGCAGACTGGGATAACCGGAGCACACGGTCTATTACTGGAAAGCA 89
Db 193 CAACATGACACCTCTGAAACCTTATAGTGTTCTCAAACCCAGCAGCCAAAANGA 252
; LOCATION: (1122) .. (1122)
; OTHER INFORMATION: n equals a,t,g, or c
; US-10-106-698-360

Query Match Score 329.6; DB 15; Length 1233;
Best Local Similarity 64.3%; Pred. No. 2.4e-95;
Matches 494; Conservative 27; Mismatches 0; Indels 0; Gaps 0;

Qy 30 GACCATGTCAGGGCATTCTGCAGACTGGGATAACCGGAGCACACGGTCTATTACTGGAAAGCA 89
Db 210 CATGAGAACAGGGCATTCTGCAGACTGGGATAACCGGAGCACACGGTCTATTACTGGAAAGCA 269
; LOCATION: (1122) .. (1122)
; OTHER INFORMATION: n equals a,t,g, or c
; US-10-106-698-360

Query Match Score 329.6; DB 15; Length 1233;
Best Local Similarity 64.3%; Pred. No. 2.4e-95;
Matches 494; Conservative 27; Mismatches 0; Indels 0; Gaps 0;

Qy 30 GACCATGTCAGGGCATTCTGCAGACTGGGATAACCGGAGCACACGGTCTATTACTGGAAAGCA 89
Db 253 AATPATCATGTCAGGGCATTCTGCAGACTGGGATAACCGGAGCACACGGTCTATTACTGGAAAGCA 312
; LOCATION: (1122) .. (1122)
; OTHER INFORMATION: n equals a,t,g, or c
; US-10-106-698-360

Query Match Score 329.6; DB 15; Length 1233;
Best Local Similarity 64.3%; Pred. No. 2.4e-95;
Matches 494; Conservative 27; Mismatches 0; Indels 0; Gaps 0;

Qy 30 GACCATGTCAGGGCATTCTGCAGACTGGGATAACCGGAGCACACGGTCTATTACTGGAAAGCA 89
Db 270 TCTCGGTGCTGCTGGCTGCTGGGAGTTACAGGTTACAGGTTACAGGTTACAGGTCAGGG 329
; LOCATION: (1122) .. (1122)
; OTHER INFORMATION: n equals a,t,g, or c
; US-10-106-698-360

| | | | | |
|--|-----|---|-----|---|
| Db | 184 | GCTGTATCCATTGCCAATAACAGTCCCCCTGGAAATAACAGTCAACTGAAAC | 243 | Best Local Similarity Matches 494; Conservative 0; Mismatches 274; Indels 0; Gaps 0; |
| Qy | 150 | GAATATGACTCTCCCGACCCTTAAGTATCAACTGACCCAAAGCTCGTAAAT | 209 | Qy 30 GACCATGTCGAGGCTCAGCTGGGATACGGCAGCAACCGTCTATTCACTGGAGGA 89 |
| Db | 244 | CAACATACAACCTCTCGAAACCTATAGTGTCTCCACACGCCAAGA | 303 | Db 1218 GATAATGGCAAGTCAGACTGGGATATGATGACAAAAATGTGAGCAA 1277 |
| Qy | 210 | CATCAGGZACAGGGCCTTCCCTCACTGTTGACTTGTGACACAGAAGAACATTAGT | 269 | Qy 90 ATTTTCCCTATTGCTGATGGTATCAACCTCAATTGAGATAAAACAAAGAGT 149 |
| Db | 304 | AATTATCATGTTGGGCACTTCCTCCATTGTAATTTCAGGACATAACGATAGT | 363 | Db 1278 GCTGATCCATGCCATGGAAATAACCGTCCTGGTATATAAACCGTAAAC 1337 |
| Qy | 270 | TCTGCGTGTGTGCTCTCACTGGAAAGTACAGGTAGCCAGGTTAACCTTCAGCGG | 329 | Qy 150 GAAATATGACTCTCCCGACCCTTAGTATCAAGTATGACCAAGCTCAGTAAAT 209 |
| Db | 364 | GCTGAAGGTGNCCTTCTCTGACAGTACAGGCTCTTCAGTCACTGGG | 423 | Db 1338 CAAACATGACACCTCTCTGAAACCTATTAGTCTCCACACGCCACGCCAAGA 1397 |
| Qy | 330 | GTCGGCTGATGACCCGGCTCCAGCACATACTGAGTGGCTAGCTGAGGCT | 389 | Qy 210 CATAGGZACAGGGCCTTCCCTCACTGTCAGCTTGACTTGTGACACAGAACAAATGAGT 269 |
| Db | 424 | CACTACAATGACATGGTCAAGACATACTGTTGATGAGTCAAATTCTGCCGAGT | 483 | Db 1398 AATTTATCATGTCGGGATTCTTCCATGTAATTGTTGAGGCAACGATAACCGTCACTGAGT 1457 |
| Qy | 390 | CCATGTTTCACTGGAAATTCAAGAACATACCCAGCTTGTGAGCCAGGCTCATGAAAC | 449 | Qy 270 TCTGGTGTGGCTCTCACTGAAAGCTCAGTGGTACGGGAGGTTACCTICACTGGG 329 |
| Db | 484 | TCACTGACTCTCACTGGAAATTCTGAAACTACCCAGCTTGTGAGGCTCAAGGC | 543 | Db 1458 GCTGAAAGTGGCTCTCTCTGACAGCTACGGCTCUTTCATTTCATCTGGG 1517 |
| Qy | 450 | AGATGGACTGGCTGCTCTGGAGTGTGTTTACAGATGGTGAACCTAAATTCCCACAGCA | 509 | Qy 330 GTCGGCTGATGACCCGGCTCCGGACACATAGTAGTGGAGTGTGAGCTGAGACT 389 |
| Db | 544 | TGTGGTTGGCAGTTATGGTGTGGAGGTGTGAGGGTTGTGAGGTCACTGAGT | 603 | Db 1518 CAGTACAAAATGACCACTGGTCAAGAACATACAGTGGATGGTCAATATTCTGCGAGCT 1577 |
| Qy | 510 | AAAGATTATGACACTTGGATTCATTAAAGAAAAGGTAACAAACTCGATTCAA | 569 | Qy 390 CCTATGTTTCACTGGATTCTGAGCAAAATAACCCAGCTTGTGAGCTCATGAAAC 449 |
| Db | 604 | GAAGATTACTTGTGCCCCCTCAACTCCCTCRAAGGAAATTAAACCAAGGCCAA | 663 | Db 1578 TCACTGACTCTCACTGGAAATTCTCAAAACTACTCAGCTTGTGAGGTGCTCAAGGC 1637 |
| Qy | 570 | TCTTGACCTATTGCTCTCCACCATCCTGGACTACTGGACATACTCCTGGTCCT | 629 | Qy 450 AGATGGACTCTGGCTCTGGAGTGTGTTTACAGATGGTGAACCTAAATTCCAACTGCA 509 |
| Db | 664 | TCTTGACCCCTCAACTCCCTCRAAGGAAATTAAACCAAGGCCAAACGAGGCCAA | 723 | Db 1638 TGATGGTTGGAGTTATGGTTTGTGAGGTRGTGTGAGGCCAACCCAAAGCTCA 1697 |
| Qy | 630 | TACAGTTCCACCTCTCTGTGAGACTCACATGGATGTTTAAAGCAACCTATAAACAT | 689 | Qy 510 AAAGATTACTGACACTTGGATCTICATTAAGAAAAGGTAAACAAAATCGATTCAAAA 569 |
| Db | 724 | GACTCATCTCCCTTATATGAGACTCACATGGTCAACTGGTCACTGTAAAGGAGGATCAAGT | 783 | Db 1698 GAAGTACTTGTATGCCCTCAAGCAATTAAACCAAGGCCAAAGAGGCCCAATTACAA 1757 |
| Qy | 690 | CAGTCTCAACAGTGGCCAATTTCGAGTCTCTGTGACAGCGGGGTTGAAGGZGC | 749 | Qy 570 TTGGACCTATTGCTCTGCTCCACATCCTGGACTACTGGACATATCCCTGGTTCTCT 629 |
| Db | 784 | CAGTCAAGGAGCTGGCAATTCCGAGCTTCTATCAAATGTTGAGGTGATAAGGC | 843 | Db 1758 TTGGACCCCTCTACTCCCTTCATCCCTGGATTCTGCTTCCTGCTCT 1817 |
| Qy | 750 | AGCTTTCTGTGATAGATGTCACTCTGTCACCCAGGGCTGGAGGGCAG | 797 | Qy 630 TACAGTTCCACCTCTCTGTGAGACTCACATGGTCAACTGGTCACTGAGT 689 |
| Db | 844 | TGTCCCCCATGAGCACACAAACCGCCAACTCTGAAGGGCAG | 891 | Db 1818 GACTCATCTCCCTCTACTCCCTTCATCCCTGGATTCTGCTTCCTGCTCT 1877 |
| RESULT 1.2 | | | | |
| US-10-158-646-47 | | | | |
| Sequence 47, Application US/10158646 | | | | |
| Publication No. US2003007105A1 | | | | |
| GENERAL INFORMATION: | | | | |
| PRIORITY FILING DATE: 2001-05-31 | | | | |
| NUMBER OF SEQ ID NOS: 78 | | | | |
| SOFTWARE: PERL Program | | | | |
| SEQ ID NO 47 | | | | |
| LENGTH: 2826 | | | | |
| TYPE: DNA | | | | |
| ORGANISM: Homo sapiens | | | | |
| FEATURE: | | | | |
| NAME/KEY: misc feature | | | | |
| OTHER INFORMATION: Incyte ID No. US20030073105A1 2395688.5 | | | | |
| US-10-158-646-47 | | | | |
| Query Match | | | | |
| 36.2%; Score 329.6; DB 15; Length 2826; | | | | |
| RESULT 1.3 | | | | |
| US-09-917-800A-1675 | | | | |
| Sequence 1675, Application US/09917800A | | | | |
| Patent No. US20020119462A1 | | | | |
| GENERAL INFORMATION: | | | | |
| APPLICANT: Mandrick, Donna | | | | |
| APPLICANT: Porter, Mark | | | | |
| APPLICANT: Johnson, Kory | | | | |
| APPLICANT: Castle, Arthur | | | | |
| APPLICANT: Elashoff, Michael | | | | |
| APPLICANT: Gene Logic, Inc. | | | | |
| TITLE OF INVENTION: Molecular Toxicology Modeling | | | | |
| FILE REFERENCE: 44921-5038-US | | | | |
| CURRENT APPLICATION NUMBER: US/09917800A | | | | |
| CURRENT FILING DATE: 2001-07-31 | | | | |
| PRIOR APPLICATION NUMBER: US 60/222,040 | | | | |
| PRIOR FILING DATE: 2000-07-31 | | | | |

PRIOR APPLICATION NUMBER: US 60/222,880
 PRIOR FILING DATE: 2000-11-02
 PRIOR APPLICATION NUMBER: US 60/290,029
 PRIOR FILING DATE: 2001-05-11
 PRIOR APPLICATION NUMBER: US 60/290,645
 PRIOR FILING DATE: 2001-05-15
 PRIOR APPLICATION NUMBER: US 60/292,336
 PRIOR FILING DATE: 2001-05-22
 PRIOR APPLICATION NUMBER: US 60/295,798
 PRIOR FILING DATE: 2001-06-06
 PRIOR APPLICATION NUMBER: US 60/297,457
 SEQ ID NO 1675
 LENGTH: 1459
 TYPE: DNA
 ORGANISM: Rattus norvegicus
 OTHER INFORMATION: Genbank Accession No. US20020119462A1 NM_019291
 US -09-917-800A-1675

Query Match 31.8%; Score 289.4; DB 9; Length 1459;
 Best Local Similarity 62.4%; Pred. No. 2.8e-82;
 Matches 471; Conservative 0; Mismatches 281; Indels 3; Gaps 1;

Qy 43 CTCAGCTGGATACCGGCAACGGTCTATTCACTGGAAATTTCCTATT 102
 Db 15 CACACTGGGATACGGAAGAACGGCAAGGAACTGGCAAGGAGTCCCCATT 74
 Qy 103 GCTGATGGTGTACGGCATCTCCATTGGATAAACCAAGAAGTGAAATAATGACTCT 162
 Db 75 GCAAATGGAGACAGTCCCCTGTGACATGCACTGGCATGACCTT 134
 Qy 163 TCCCTCGACGACACTTAGTATCAGTAAATCTCAGAACAGC 222
 Db 135 TCCCTACAGCTCTGCTCATATGTACATTAAGTTGCCAGAGCTATGTCAACAT 194
 Qy 223 GGCGATTCCTCAATGTTGACAGACAGAAATAATCAGTTGCGTGGTGT 282
 Db 195 GGCATATTCCTCAACGTGTTGAGTTGATGCTTCCAGGACTTGCAGTGTAAAGAGGGA 254
 Qy 283 CCTCTCACTGGAGCTACGGTACGGGAGGTACCCCTCACTGGGGTCGGCTGATGAC 342
 Db 255 CCCCTAGTGGCTCTAACAGTGTACCTGGGTCATCTGATEGC 314
 Qy 343 CACGGCTCGAGCACATGAGTGGAGTGAGCTGAGTGTGGTGTAC 402
 Db 315 CAGGGCTCTGAGCACACCGTGAACAAAAAAATATGCTGAGCTTCACTGGTTCAC 374
 Qy 403 TGGATTGAGCAATAACCCACGCTTGTGAGCAGCTCATGACCAAGTGGACTGGGT 462
 Db 375 TGGAAACAC--CAGATGGGATTGTTGGAGACCCAGATGGACTGGT 431
 Qy 463 GTCTTGGAGTGTGTTTACAGATTGGTAACTGCAAAGAATACTGAC 522
 Db 432 GTTTGGTGTATTTTGAATTGGACCTGGCTTACAGGCCCTTACGAAATCACTGAA 491
 Qy 523 ACTTGGATTCCATTAAAGAAGGGTAAACACTCGATCAAAATTGACCTAATG 582
 Db 492 GCACTGGATTCCATTAAACAAAGGGAAACCTGGCTGAGCTTGTCACTGGC 551
 Qy 583 TCTCTGCTTCACATCCTGGGACTACTGGACATATCCTGGTCTCTPACAGTCCACT 642
 Db 552 TCCCTCTCTGCAACTTGGACTACTGGACATATCCTGGCTCAGACCACTGGCC 611
 Qy 643 CTTCCTGAGAGCTCACAGTGTGTTAACAACTTAACATGAGCTCAAG 702
 Db 612 CTGCTGGAAATGTTGACCTGGGACTGAGTGGTCAAGAAACCTTACAGTCAAG 671

RESULT 14
 US 10-191-803-124
 ; Sequence 124, Application US/10191803
 ; Publication No. US20040014040A1
 ; GENERAL INFORMATION:
 ; APPLICANT: MENDRICK, Donna
 ; APPLICANT: PORTER, Mark
 ; APPLICANT: JOHNSON, Kory
 ; APPLICANT: ELASHOFF, Michael
 ; APPLICANT: CASTLE, Arthur
 ; APPLICANT: HIGGS, Brandon
 ; TITLE OF INVENTION: Cardiotoxin Molecular Toxicology Modeling
 ; FILE REFERENCE: 44921-5090US
 ; CURRENT APPLICATION NUMBER: US/10/191,803
 ; CURRENT FILING DATE: 2002-07-10
 ; PRIOR APPLICATION NUMBER: US 60/303,819
 ; PRIOR FILING DATE: 2001-07-10
 ; PRIOR APPLICATION NUMBER: US 60/305,623
 ; PRIOR FILING DATE: 2001-07-17
 ; PRIOR APPLICATION NUMBER: US 60/369,351
 ; PRIOR FILING DATE: 2002-04-03
 ; PRIOR APPLICATION NUMBER: US 60/377,611
 ; PRIOR FILING DATE: 2002-05-06
 ; NUMBER OF SEQ ID NOS: 1140
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO: 1459
 ; LENGTH: 1459
 ; TYPE: DNA
 ; ORGANISM: Rattus norvegicus
 ; FEATURE:
 ; OTHER INFORMATION: Genbank Accession No. US20040014040A1 NM_019291
 US-10-191-803-124

Query Match 31.8%; Score 289.4; DB 16; Length 1459;
 Best Local Similarity 62.4%; Pred. No. 2.8e-82;
 Matches 471; Conservative 0; Mismatches 281; Indels 3; Gaps 1;

Qy 43 CTAGCTGGGATAACCGGACAACGGTCTTATCACTGGAGTATTGGGATTTCCTATT 102
 Db 15 CACACTGGGATACGGAAGAACGGCAAGGAACTGGCAAGGAGTCCCCATT 74
 Qy 103 GCTGATGGTGTACGGCATCTCCATTGGATAAACCAAGAAGTGAAATAATGACTCT 162
 Db 75 GCAAATGGAGACAGTCCCCTGTGACATGCACTGGCATGACCTT 134
 Qy 163 TCCCTCGACGACACTTAGTATCAGTAAATCTCAGAACAGC 222
 Db 135 TCCCTACAGCTCTGCTCATATGTACATTAAGTTGCCAGAGCTATGTCAACAT 194
 Qy 223 GGCGATTCCTCAATGTTGACAGACAGAAATAATCAGTTGCGTGGTGT 282
 Db 195 GGCATATTCCTCAACAGTGTACCTGGGTCATCTGATEGC 314
 Qy 283 CCTCTCACTGGAGCTACGGTACGGGAGGTACCCCTCACTGGGGTCGGCTGATGAC 342
 Db 255 CCCCTAGTGGCTCTAACAGTGTACCTGGGTCATCTGATEGC 314
 Qy 343 CACGGCTCGAGCACATGAGTGGAGTGAGCTGAGTGTGGTGTAC 402
 Db 315 CAGGGCTCTGAGCACACCGTGAACAAAAAAATATGCTGAGCTTCACTGGTTCAC 374
 Qy 403 TGGATTGAGCAATAACCCACGCTTGTGAGCAGCTCATGACCAAGTGGACTGGGT 462
 Db 375 TGGAAACAC--CAGATGGGATTGTTGGAGACCCAGATGGACTGGT 431
 Qy 463 GTCTTGGAGTGTGTTTACAGATTGGTAACTGCAAAGAATACTGAC 522
 Db 432 GTTTGGTGTATTTTGAATTGGACCTGGCTTACAGGCCCTTACGAAATCACTGAA 491
 Qy 523 ACTTGGATTCCATTAAAGAAGGGTAAACACTCGATCAAAATTGACCTAATG 582
 Db 492 GCACTGGATTCCATTAAACAAAGGGAAACCTGGCTGAGCTTGTCACTGGC 551
 Qy 583 TCTCTGCTTCACATCCTGGGACTACTGGACATATCCTGGTCTCTPACAGTCCACT 642
 Db 552 TCCCTCTCTGCAACTTGGACTACTGGACATATCCTGGCTCAGACCACTGGCC 611
 Qy 643 CTTCCTGAGAGCTCACAGTGTGTTAACAACTTAACATGAGCTCAAG 702
 Db 612 CTGCTGGAAATGTTGACCTGGGACTGAGTGGTCAAGAAACCTTACAGTCAAG 671

| | | |
|----|-----|--|
| Db | 375 | TGGAAACAG --- CAAATAGGGATTGAAAAGGTGTGAGCACCCAGATGGCTTGGCT 431 |
| Qy | 463 | GTCCTGGAGTTTACAGATTTGTGAACTTAATTCCCCAACTGCAAAAGATTACTGAC 522 |
| Db | 432 | GTTTGGATTTTTGAGATTTGACCTGCTACAAGGCCCTACGAAATCATCTGAA 491 |
| Qy | 523 | ACTTTGGATTCCATTAAAGGTTAACAAACTCGATTCACAAATTGACTTATG 582 |
| Db | 492 | GCACTGCATTCCATTAAACAGGGAAACGTGAGCCTTTGTAATTGATCTTGC 551 |
| Qy | 583 | TCTCTGGPTTCCACCATCCCTGGACTACTGGACATACTCTGGTTCTTACAGTCCACT 642 |
| Db | 552 | TCCCTTCTCCCTGAAACCTGGACTACTGACATACCTGGCTCTGACCATCTGCC 611 |
| Qy | 643 | CCTCTTGAGAGTGTGACATGGATTGTTTAAAGCAACCTATAAACATCAGCTCTAACAG 702 |
| Db | 612 | CTGTGGATGTGACCTCTGGATACTGTCTCAAGGAAACCATTACTGTCAGGTGAGCAG 671 |
| Qy | 703 | CTGGCCAATTTCGAGTCCTGGTCACTGGAGCTTCTGTGA 762 |
| Db | 672 | ATGTCATTTCCGTAACACTGAACTTCATTGGGGGGCTGAAGAACTGGATGCTG 731 |
| Qy | 763 | TAGAGTCACATCTGTCACCCAGCTGGAGCCAG 797 |
| Db | 732 | GACACTGGGTCCAGCTCACGGCTGAAGACAG 766 |

RESULT 15 JS-09-988-292-17

Sequence 17 - Application US/09988292
Publication No. US20020086314A1
GENERAL INFORMATION:

APPLICANT: Yu, Guo-Liang
Rosen, Craig
TITLE OF INVENTION: Colon Specific Genes and Proteins
NUMBER OF SEQUENCES: 24
CORRESPONDENCE ADDRESS:
ADDRESSEE: Carella, Byrne, Bain, Gilfillan, Cecc

STREET: 6 Becker Farm Road
CITY: Roseland
Stewart & Olstein

STATE: NJ
COUNTRY: USA

ZIP: 07068-1739
UTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPATIBILITY: IBM PC compatible

COMPUTER: IBM PC COMPATIBLE
OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

ENT APPLICATION DATA: _____

APPLICATION NUMBER: US/09/988,292
FILING DATE: 10 Nov 2001

FILING DATE: 19-N8V-2001
CLASSIFICATION: Unknown

CLASSIFICATION: UNCLASSIFIED

APPLICATION NUMBER: 09/224,110

FILING DATE: <Unknown>

NAME: Ferraro, Gregory D
ATTORNEY/AGENT INFORMATION:

NAME: FERRARI, GREGORY D.
REGISTRATION NUMBER: 36-1134

REFERENCE/DOCKET NUMBER: 325800-435
TELECOMMUNICATION INFORMATION:

TELEPHONE: 201-994-1700

TELEFAX: 201-994-1744

INFORMATION FOR SEQ ID NO: 17:
SEQUENCE CHARACTERISTICS:

SEQUENCE CHARACTERISTICS: LENGTH: 605 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear
WAVELENGTH: 532 nm

Job time : 502 secs

GenCore version 5.1.6
 Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 13, 2004, 20:55:55 ; Search time 424 Seconds
 (without alignments)
 9127.614 Million cell updates/sec

Title: US-10-069-434-4
 Perfect score: 911
 Sequence: 1 cggaaatccggctcgaggatcc.....coccaatgccaggggaaaa 911

Scoring table: IDENTITY_NUC Gapop 10_0 , Gapext 1.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0
 Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
 Maximum Match 100%
 Listing first 45 summaries

| Database : | N_Geneseq_29Jan04:* |
|------------|---------------------|
| | 1: geneseqn1900s:* |
| | 2: geneseqn1900s:* |
| | 3: geneseqn2000s:* |
| | 4: geneseqn2001as:* |
| | 5: geneseqn2001bs:* |
| | 6: geneseqn2002s:* |
| | 7: geneseqn2003as:* |
| | 8: geneseqn2003bs:* |
| | 9: geneseqn2003cs:* |
| | 10: geneseqn2004s:* |

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query | Match Length | DB ID | Description |
|------------|-------|-------|--------------|-----------|---------------------|
| 1 | 911 | 100.0 | 911 | ABA9791 | Ab9791 Human lya |
| 2 | 884.8 | 97.1 | 2094 | ABA34690 | Abx34690 Human mdd |
| 3 | 759 | 83.3 | 1248 | Aas30989 | Aas30989 Human dia |
| 4 | 745.8 | 81.9 | 1023 | Ada41388 | Aad41188 Human NZM |
| 5 | 740.6 | 81.3 | 828 | Abt33350 | Abt33350 NOVX DNA |
| 6 | 728.8 | 80.0 | 833 | Abt33351 | Abt33351 NOVX DNA |
| 7 | 493.2 | 54.1 | 688 | Aaf22342 | Aaf22342 Human sec |
| c 8 | 444.4 | 48.8 | 462 | AdS68431 | Aas68431 DNA encod |
| 9 | 444.4 | 48.8 | 496 | ACH26864 | Ach26864 Human adu |
| 10 | 421 | 46.2 | 462 | AAI192460 | Aai192460 Human pol |
| 11 | 331.2 | 36.4 | 2785 | AAI43637 | Aai43637 Human car |
| 12 | 329.6 | 36.2 | 1233 | AAH33294 | Aah33294 Human col |
| 13 | 329.6 | 36.2 | 1244 | ADA10961 | Ada10961 Human CDN |
| 14 | 329.6 | 36.2 | 2826 | ABD59160 | Abd59160 Human car |
| 15 | 289.4 | 31.8 | 1459 | ABE63768 | Abk63768 Rat sequ |
| 16 | 289.4 | 31.8 | 1459 | ABD58362 | Abd58362 Toxicity- |
| 17 | 289.4 | 31.3 | 605 | AAI45889 | Aat45889 Human col |
| 18 | 284.8 | 31.3 | 605 | AAV16677 | Aav16677 Polynucle |
| 19 | 284.8 | 31.3 | 605 | ABK15465 | Abk15465 Human col |
| 20 | 284.8 | 31.3 | 605 | AB52546 | Abs52546 Human col |
| 21 | 280 | 30.7 | 783 | AAV47615 | Aav47615 Nucleotid |
| 22 | 280 | 30.7 | 864 | AAU73088 | Aat73088 DNA for f |
| 23 | 280 | 30.7 | 1759 | ADA10998 | Ada10998 Human CDN |

ALIGNMENTS

| RESULT 1 | ID | ABA9791 | ABA9791 standard; DNA; 911 BP. |
|----------|-----|---------|--|
| | XX | | |
| | AC | | |
| | XX | | |
| | DT | | 18-JUN-2002 (First entry) |
| | DE | | Human lyase HLVA-1 encoding cDNA. |
| | XX | | |
| | KW | | Cytostatic; anticonvulsant; cerebroprotective; nootropic; virucide; |
| | KW | | neuroprotective; antibacterial; antidiabetic; antiinflammatory; antigout; |
| | KW | | ophthalmological; hypotensive; immunosuppressive; dermatological; |
| | KW | | nephrotrophic; antichyroid; thyromimetic; osteopathic; antiparasitic; |
| | KW | | antiluler; fungicide; antiparasitic; tricyclic; protozoicide; tranquilliser; cancer; |
| | KW | | neuroleptic; diagnosis; treatment; immunological disorder; AIDS; allergy; |
| | KW | | acquired immunodeficiency syndrome; asthma; HIVA; infection; anaemia; |
| | KW | | Crohn's disease; multiple sclerosis; atherosclerosis; osteoporosis; |
| | KW | | rheumatoid arthritis; Alzheimer's; Parkinson's disease; epilepsy; stroke; |
| | KW | | muscular dystrophy; Down's syndrome; myasthenia gravis; glaucoma; |
| | XX | | transgenic; gene therapy; drug screening; human lyase; gene; ss; |
| | XX | | Homo sapiens. |
| | OS | | |
| | XX | | |
| | PH | | Location/Qualifiers |
| | CD5 | | |
| | FT | | 34 .762 |
| | PT | | |
| | PT | | /product= "HLVA-1" |
| | FT | | /note= "Human lyase 1" |
| | XX | | |
| | PN | | WO200200840-A2. |
| | XX | | |
| | PR | | 23-JUN-2000; 2000US-0213383P. |
| | PR | | 30-JUN-2000; 2000US-0215544P. |
| | PR | | 04-AUG-2000; 2000US-022218P. |
| | XX | | (INCYT-) INCYTE GENOMICS INC. |
| | XX | | Thornton M, Ramkumar J, Tribouley CM, Yue H, Nguyen DB, Yao MG; |
| | PI | | Patterson C, Gandhi AR, |
| | PI | | Burford N, Thangavelu K, Baughn MR; |
| | XX | | WPI; 2002-139910/18. |

| | | | |
|----|---|---|--|
| DR | P-PSDB; ABB08900. | Db | 481 CAGATTCGTGAACTTAATGCCAACAAAGATACTGACATTGGATTCCATTAA 540 |
| XX | New isolated human lyase polypeptide for diagnosing, treating and preventing e.g. glaucoma, ocular hypertension, stroke, asthma, or gout. | QY | 541 GAAAAGGTAAACAAACTCGATTCAAAATTGACTTATGCTTCACCATCC 600 |
| PT | | DB | 541 GAAAAGGTAAACAAACTCGATTCAAAATTGACTTATGCTTCACCATCC 600 |
| XX | | QY | 601 TCGGACTACTGGACATACTCTGTCTCTTAACTGTTCAACCTCTCTTGAGGTGTCACA 660 |
| PS | Claim 11: Page 99-100; 101PP; English. | DB | 601 TCGGACTACTGGACATACTCTGTCTCTTAACTGTTCAACCTCTCTTGAGGTGTCACA 660 |
| CC | The present sequence represents a cDNA encoding a human lyase polypeptide (HLYA) given in ABB08900. The specification describes an isolated HLYA polypeptide or a nucleic acid that encodes it. The invention has cytostatic, nootropic, anticonvulsant, cerebroprotective, virucide, neuroprotective, anti-HIV, anti-parkinsonian, antiparasitic, antidiabetic, antiinflammatory, ophthalmological, hypotensive, antiallergic, antiulcer, antianemic, antiasthmatic, antiatherosclerotic, immunosuppressive, dermatological, nephrological, osteopathic, antipsoriatic, antirheumatic, antiarthritic, dermatological, fungicide, antiparasitic, protozoicide, tranquilliser and neuroleptic applications. The protein of the invention may be used to screen for potential HLYA agonists or antagonists; detect the presence of HLYA associated disorders; assess the toxicity of a test compound. The HLYA proteins and polynucleotides are useful in diagnosis, treatment and prevention of immunological disorders e.g. AIDS, allergy, anaemia, asthma, infection, Crohn's disease, multiple sclerosis, atherosclerosis, rheumatoid arthritis, osteoporosis; cancer; neurological disorder e.g. Alzheimer's and Parkinson's disease, epilepsy, stroke, muscular dystrophy, Down's syndrome, myasthenia gravis; glaucoma. HLYA polynucleotides are used for creating humanised/transgenic animals to model human diseases; somatic or germline gene therapy; for generating hybridisation probes for e.g. gene mapping; detecting differences in chromosomal location due to e.g. translocation; generating a transcript image of a tissue/cell type. Antibodies which bind to the HLYA proteins are used for diagnosis of HLYA associated disorders or monitoring patients being treated with HLYA or agonists, antagonists or inhibitors | QY | 661 TGGATGTTTAAAGGAAACCTATAAACATCAGCTCACACGTCGCGCAAATTTCGAGT 720 |
| CC | | DB | 661 TGGATGTTTAAAGGAAACCTATAAACATCAGCTCACACGTCGCGCAAATTTCGAGT 720 |
| CC | | QY | 721 CTCCTGGCACCGGGGGTCAAGNGCAGGTTCTGTGATAGAGTCCTGACTCTCA 780 |
| CC | | DB | 721 CTCCTGGCACAGGGGGTCAAGGAGGAACTTATGATAGAGTCCTGACTCTCA 780 |
| CC | | QY | 781 CCCAGGTGTCCTCCACCTCAGGCTCCAGTGGTCAATCTGGTATATGAGCTTCCAACTCTGTGACTCA 840 |
| CC | | DB | 781 CCCAGGTGTCCTCCACCTCAGGCTCCAGTGGTCAATCTGGTATATGAGCTTCCAACTCTGTGACTCA 840 |
| CC | | QY | 841 AGTGAATCTCTCCACCTCAGGCTCCAGTGGTCAATCTGGTATATGAGCTTCCAAATGC 900 |
| CC | | DB | 841 AGTGAATCTCTCCACCTCAGGCTCCAGTGGTCAATCTGGTATATGAGCTTCCAAATGC 900 |
| CC | Sequence 911 BP; 238 A; 235 C; 205 G; 233 T; 0 U; 0 Other; | QY | 901 CACGGGAAAA 911 |
| CC | Best Local Similarity 100.0% ; Score 911; DB 6; Length 911; | DB | 901 CACGGGAAAA 911 |
| CC | Matches 911; Conservative 0; Mismatches 0; Indels 0; Gaps 0; | QY | 901 CACGGGAAAA 911 |
| CC | Score 911; Pred. No. 3.7e-206; ID ABX34690 standard; cDNA; 2094 BP. | DB | ABX34690; AC |
| CC | XX | XX | ABX34690; |
| CC | XX | XX | DT 13-FEB-2003 (first entry) |
| CC | XX | XX | Human mddt cDNA SEQ ID 251. |
| CC | XX | XX | MDDT; human; disease detection and treatment molecule polypeptide; anti-inflammatory; immunosuppressive; osteopathic; cytostatic; anti-HIV; KW nephrotropic; antipsoriatic; hepatotropic; haemostatic; gene therapy; protein replacement therapy; cell proliferative disorder; KW adenocarcinoma; leukaemia; melanoma; sarcoma; anemia; Crohn's disease; acquired immunodeficiency syndrome; AIDS; KW Goodpasture's syndrome; inflammation; osteoporosis; thrombocytopaenia; psoriasis; hepatitis; gene; ss. |
| CC | XX | XX | Homosapiens |
| CC | XX | XX | OS Homo sapiens |
| CC | XX | XX | PF 27-MAR-2002; 2002NO-US00944. |
| CC | XX | XX | PR 28-MAR-2001; 2001US-0279619P. |
| CC | XX | XX | PR 29-MAR-2001; 2001US-0200067P. |
| CC | XX | XX | PR 29-MAR-2001; 2001US-0200068P. |
| CC | XX | XX | PR 16-MAY-2001; 2001US-0211280P. |
| CC | XX | XX | PR 17-MAY-2001; 2001US-0211829P. |
| CC | XX | XX | PR 17-MAY-2001; 2001US-0211849P. |
| CC | XX | XX | PR 19-JUN-2001; 2001US-0239428P. |
| CC | XX | XX | PR 20-JUN-2001; 2001US-0239776P. |
| CC | XX | XX | PR 20-JUN-2001; 2001US-0300001P. |
| CC | XX | XX | PA (INCY-) INCYTE GENOMICS INC. |
| CC | XX | XX | PI Daffeo A, Jones AL, Tran AB, Dahl CR, Gietzen D, Chinn J, DuFour GB, Hillman JR, Yu JY, Tuason O, Yap PE, Amshey SR; |
| CC | QY | 481 CAGATGGTGAACCTTAATCCAACTGACATACTGGATTCCATTAA 540 | |
| CC | DB | 481 CCCAGCTTGTGAGGCACTCATGAAACAGATGGACCTGCTTGGAGGTGTTTA 480 | |
| CC | DB | 481 CACACTTGAGGCAAGTCAAGCCAGATGACTGCTTGGAGGTGTTTA 480 | |

| | | | |
|----|---|----|---|
| PI | Daugherty SC, Dam TC, Liu TF, Nguyen DA, Kleefeld Y, Gerstein EH; | Qy | 541 GAAAAGGTAAACAAACTCGATTCAAATTTGACCTATTGCTCTGCTTCACCATCC 600 |
| PI | Peralta CH, David MH, Lewis SA, Chen AJ, Panzer SR, Harris B; | Db | 555 GAAAAGGTAAACAAACTCGATTCAAATTTGACCTATTGCTCTGCTTCACCATCC 614 |
| XX | | | |
| DR | WPI: 2003-058431/05. | Qy | 601 TGGGACTACTGGACATAATCCGGTTCTTACAGTCCACCTCTTGAGAGTGTACA 660 |
| XX | New purified disease detection and treatment molecule proteins and | Db | 615 TGGGACTACTGGACATAATCCGGTTCTTACAGTCCACCTCTTGAGAGTGTACA 674 |
| PT | polypeptides, useful for diagnosing, treating or preventing cancers | Qy | 661 TGGATTGTTAACGGACCTATAAACATCAGCTCAACAGTGCCTAACATTTCAGT 720 |
| PT | (e.g. leukemia or sarcoma), anemia, Crohn's disease, AIDS, osteoporosis | Db | 675 TGGATTGTTAACGGACCTATAAACATCAGCTCAACAGTGCCTAACATTTCAGT 734 |
| XX | or hepatitis. | | |
| PS | Claim 1; SEQ ID NO 251; 339pp + Sequence Listing; English. | Qy | 721 CTCCCTGGCAGCGGAGGGTGAAGGCCAGCACTTTCTGTGATAGAGTCAGCTGTCA 780 |
| XX | This invention describes a novel disease detection and treatment molecule | Db | 735 CTCCTGGCAGCGGAGGGTGAAGGCCAGCACTTTCTGTGATAGAGTCAGCTGTCA 794 |
| CC | polypeptide (MDPT) which has anti-inflammatory, immunosuppressive, | Qy | 781 CCCAGGCTGGGGGGGGGTGTTACAACTTGCTTAATGAGCCCTCAACTCTGGACTCA 840 |
| CC | osteopatric, cytostatic, anti-HIV, haemostatic, nephrotropic, | Db | 795 CCCAGGCTGGGGGGGG-GGTACAACTCTGGCTTAATGAGCCCTCAACTCTGGACTCA 853 |
| CC | antihaemiac, antipsoriatic and hepatotropic activity. The polynucleotides | Qy | 841 AGTGAATCTCCACCTAGCCCTCCAGAGTCTGACCACTGGCATGACTTTCCDAATG 899 |
| CC | and the polypeptides of the invention can be used for gene therapy, | Db | 854 AGTGAATCTCCACCTAGCCCTCCAGCTGACCACTGGCATGACTTTCCDAATG 913 |
| CC | protein replacement therapy and are useful for treating a variety of | Qy | 900 CCAGGGGGAAAA 911 |
| CC | diseases or conditions. These polypeptides or polynucleotides are | Db | 914 CCAAGGGGAAA 925 |
| CC | particularly useful for diagnosing, treating or preventing cell | | |
| CC | proliferative disorders (e.g. cancers including adenocarcinoma, | | |
| CC | leukaemia, lymphoma, melanoma, myeloma or sarcoma), anaemia, Crohn's | | |
| CC | disease, acquired immunodeficiency syndrome (AIDS), Goodpasture's | | |
| CC | syndrome, inflammatory, osteoporosis, psoriasis, thrombocytopaenia, psoriasis or | | |
| CC | hepatitis. ABX3440-ABU1450-ABU1845, described in the disclosure of the invention. NOTE: | | |
| CC | The sequence data for this patient did not form part of the printed | | |
| CC | specification, but was obtained in electronic format from WIPO at | | |
| CC | ftp.wipo.int/pub/published_pct_sequences | | |
| XX | Sequence 2094 BP; 555 A; 500 C; 480 G; 559 T; 0 U; 0 Other; | | RESULT 3 |
| SQ | Query Match 97.1%; Score 884.8%; DB 7; Length 2094; | | AAS30989 |
| | Best Local Similarity 99.6%; Pred. No. 2.1e-277; | | ID AAS30989 standard; cDNA; 1248 BP. |
| | Matches 908; Conservative 0; Mismatches 2; Indels 2; Gaps 2; | | XX |
| Qy | 1 CGGAATTCTGGCTGAGTTCACCCCCGAGGCCATGCGAGCTAACGCC 60 | | XX |
| Db | 15 CGGAATTCTGGCTGAGTTCACCCCCGAGGCCATGCGAGCTAACGCC 74 | | AC |
| Qy | 61 GAGCACACAGCTCTTCACTTGAAAGATTTCCTTAATGCTATGGTGATCAGCAA 120 | | XX |
| Db | 75 GAGCACACAGCTCTTCACTTGAAAGATTTCCTTAATGCTATGGTGATCAGCAA 134 | | DT |
| Qy | 121 TCCTCAATTGAGATTAAACCAAAAGAAGTGAATATGCTTCCTCCGACCTAGT 180 | | 04-DEC-2001 (first entry) |
| Db | 135 TCCTCAATTGAGATTAAACCAAAAGAAGTGAATATGCTTCCTCCGACCTAGT 194 | | |
| Qy | 181 ATCAAGTATGACCCAAGCTCAGCTAAATCATCAGCAACAGGCCATTCTTCAATGT 240 | | |
| Db | 195 ATCAAGTATGACCCAAGCTCAGCTAAATCATCAGCAACAGGCCATTCTTCAATGT 254 | | |
| Qy | 241 GACTTGTATGACACAGGAGAACAAATCAGTGTCTGCTGCTCTCTACATCTGGAAACTAC 300 | | XX |
| Db | 255 GACTTGTATGACACAGGAGAACAAATCAGTGTCTGCTGCTCTCTACATCTGGAAACTAC 314 | | PR |
| Qy | 301 AGGTAAAGGCAAGTTCACTTCACTGGGGTCCGCTGATGACCACTGGGACACATA 360 | | 24-FEB-2000; 20000US-0184693P. |
| Db | 315 AGGTAAAGGCAAGTTCACTTCACTGGGGTCCGCTGATGACCACTGGGACACATA 374 | | PR |
| Qy | 361 CTAGATGCGAGCTGATGCTGAGCTCAGCTTCACTGGAAATTGAGACAATA 420 | | 24-FEB-2000; 20000US-0184677P. |
| Db | 375 GTAGATGCGAGCTGATGCTGAGCTCAGCTTCACTGGAAATTGAGACAATA 434 | | PR |
| Qy | 421 CCCAGCTTGTGAGGAGCTGATGACCACTGGGACTGCTGCTGAGCTGTTTA 480 | | 24-FEB-2000; 20000US-0184770P. |
| Db | 435 CCCAGCTTGTGAGGAGCTGATGACCACTGGGACTGCTGCTGAGCTGTTTA 494 | | PR |
| Qy | 481 CAGATGTTGAACTTAATTCCAAACTGAAAGATTACTGACACTTGGATCCATTAA 540 | | 24-FEB-2000; 20000US-0184713P. |
| Db | 495 CAGATGTTGAACTTAATTCCAAACTGAAAGATTACTGACACTTGGATCCATTAA 554 | | PR |

| | |
|------------------------------------|---|
| CC | useful for producing non-human transgenic animals which are useful for |
| CC | studying the function and/or activity of NOVX protein and for identifying |
| CC | and/or evaluating modulators of NOVX protein activity. The NOVX nucleic |
| CC | acids can be used in gene therapy. This polynucleotide sequence |
| CC | represents a NOVX DNA sequence of the invention. |
| XX | |
| SQ | Sequence 828 BP; 224 A; 214 C; 181 G; 209 T; 0 U; 0 Other; |
| | Query Match 81-3%; Score 740.6; DB 7; Length 828; |
| | Best Local Similarity 96.9%; Pred. No. 1.4e-230; |
| | Matches 755; Conservative 0; Mi-matches 24; Indels 0; Gaps 0 |
| Qy | 19 CCACCCGGGGACATGCGGATAACCGCTATT 78 |
| Db | 1 CCACCCGGGGACATGCGGATAACCGCTATT 60 |
| Qy | 79 CACTGAAAGAATTTCCTTCAATGAGATCTCAGCATCTGATGGTCAATGAGATTAA 138 |
| Db | 61 CACTGAAAGAATTTCCTTCAATGAGATTAA 120 |
| Qy | 139 ACCAAGAACGAAATGAAATGACTCTCCCTGGACGACTTAGTACAGTACGAAAGC 198 |
| Db | 121 ACCAAGAACGAAATGAAATGACTCTCCGGACGACTTAGTACAGTACGAAAGC 180 |
| Qy | 199 TCAGCTAAATCATGAAACAGGCCATTCCTTAATGTCAGCATGACAGAG 258 |
| Db | 181 TCAGCTAAATCATGAAACAGGCCATTCCTTAATGTCAGCATGACAGAG 240 |
| Qy | 259 AACAAATCAGTCGGTCTCTCACTGGAGCTACGGTACGGTACGGTACGGTAC 318 |
| Db | 241 AACAAATCAGTCGGTCTCTCACTGGAGCTACGGTACGGTACGGTACGGTAC 300 |
| Qy | 319 CTTCACTGGGGTCGGCTGTATGACGACAGGGCTCGAGGACATAGTAGGGATCTGAGT 378 |
| Db | 301 CTTCACTGGGGTCGGCTGTATGACGACAGGGCTCGAGT 360 |
| Qy | 379 GCTGGAGGTCCATGTTGACTGGAAATACCCAGCTTGTGAGCT 348 |
| Db | 361 GCTGGAGGTCCATGTTGACTGGAAATACCCAGCTTGTGAGCT 420 |
| Qy | 439 GCTCATGAAACCAGTGGACTGGTTTACAGATTGGTCAAACCTTAAT 498 |
| Db | 421 GCTCATGAAACCAGTGGACTGGTTTACAGATTGGTCAAACCTTAAT 480 |
| Qy | 499 TCCCAACTGAAAAGATTACTGACACTTGGAATTCCATTAAAGAAAGGTAAACAACT 558 |
| Db | 481 TCCCAACTGAAAAGATTACTGACACTTGGAATTCCATTAAAGAAAGGTAAACAACT 540 |
| Qy | 569 CGATTCACAAATTGACCTATTGCTTCACCATCTGGGACTACTGGACATAT 618 |
| Db | 541 CGATTCACAAATTGACCTATTGACCTATTGCTTCACCATCTGGGACTACTGGACATAT 600 |
| Qy | 619 CCTGGTCTCTTCACTGAACTGGTCCACATGATTGTTAAAGGA 678 |
| Db | 601 CCTGGTCTCTTCACTGAACTGGTCCACATGATTGTTAAAGGA 660 |
| Qy | 679 CCTATAAACATGACGCTCAACAGTGGCCAATTTGCGACTCTCTGCAAGCGGAG 738 |
| Db | 661 CCTATAAACATGACGCTCAACAGTGGCCAATTTGCGACTCTGCAAGCGGAG 720 |
| Qy | 739 GTGTGAGCAGCAGCAGCTTCTGTGATAGAGTCTCACCTGTCAACCGCCCCACCAAGCC 797 |
| Db | 721 GTGTGAGCAGCAGCTTCTGTGATAGAGTCTCACCTGTCAACCGCCCCACCAAGCC 779 |
| RESULT 6 | |
| ABT33351 | |
| ID ABT33351 standard; DNA; 833 BP. | |
| XX | |
| XX | |
| AC | |
| XX | |
| DT 22-MAY-2003 (first entry) | |

| | | |
|-----|--|-----|
| 514 | ATTACTGACATTTGGATCCATTAAAGAAAAGGTAAACAAACTCGATTCAAAATTTC | 573 |
| 501 | ATTACTGACATTTGGATCCATTAAAGAAAAGGTAAACAAACTCGATTCAAAATTTC | 560 |
| 574 | GACCTATTGCTCTGCTTCCACCATCCTCGGAATCTGGACATPATCCGGTTCTCTTACA | 633 |
| 561 | GACCTATTGCTCTGCTTCCACCATCCTGGACTACTGGACATPATCTGGTTCTCTTACA | 620 |
| 634 | GTTCCACCTCTTGTAGAGTGTCACTGGATTGTTAAAGCACCTATAAACATCAGC | 693 |
| 621 | GTTCCACCTCTTGTAGAGTGTCACTGGATTGTTAAAGCACCTATAAACATCAGC | 680 |
| 694 | TCTCAACAGTGGCCAATTTCGZAGTCCTGCAAGGGTAAAGCAGCAGT | 753 |
| 681 | TCTCAACAGTGGCCAATTTCGZAGTCCTGCAAGGGTAAAGCAGCAGT | 740 |
| 754 | TTTCGTCGATAGACTCTCACTCTGTCACCCAGGCTGGACGGCAG | 797 |
| 741 | TTTCGTCGATAGACTCTCACTCTGTCACCCAGGCTGGACGGCAG | 784 |

RESULT 7
 AF22342
 AA22342 standard; cDNA; 688 BP.
 X
 C
 XAF22342;
 T
 26-MAR-2001 (first entry)

X Human secreted protein gene 27 SEQ ID NO:37.
 X Human; secreted protein; diagnosis; immunosuppressive; antiarthritic;
 W antirheumatic; antiproliferative; cytostatic; cardiotonic; vasoconstrictive;
 W cerebroprotective; nootropic; neuroprotective; antibacterial; viricide;
 W fungicide; opthalmological; vulvar; gene therapy; neoplasm;
 W autoimmune disease; rheumatoid arthritis; hyperproliferative disorder;
 W cardiovascular disorder; cardiac arrest; cerebrovascular disorder;
 W cerebral ischaemia; angiogenesis; nervous system disorder; infection;
 W Alzheimer's disease; ocular disorder; corneal infection; wound healing/
 W skin acne; food additive; preservative; ss.

Homo sapiens.
WO200061748-A1.

19-OCT-2000.

06-AEP-2000; 2000WU-US00890Z.
X 09-APR-1999; 99US-0128696P.
PR 14-JAN-2000; 2000US-0178069P.
X (HUMA) HUMAN GENOME SCI INC.
X Rosen CA, Ruben SM,
X Komatsoulis G;

WBI : 3000 638333 / 61

WPI; 2000-838388; 81:
P-PSDB; AAB63075;
OR

xx

New nucleic acid motif diagnosing prevention

used as food additive

Exhibit 117 Page 1

XXVII

AAF22316 to AAF22363

E8 AAB63096. AAB6309

activities based on

Examples of activities

antineurameric; amiprocycline; n-

ECC fungicide; ophthalmic

proteins can be used to prevent, treat or ameliorate a medical condition in e.g. humans, mice, rabbits, goats, horses, cats, dogs, chickens or sheep. They are also used in diagnosing a pathological condition or susceptibility to a pathological condition. Disorders which are diagnosed or treated include autoimmune diseases e.g. rheumatoid arthritis, hyperproliferative disorders e.g. neoplasms of the breast or liver, cardiovascular disorders e.g. cardiac arrest, cerebrovascular disorders e.g. cerebral ischaemia, angiogenesis, nervous system disorders e.g. Alzheimer's disease, infections caused by bacteria, viruses and fungi and ocular disorders e.g. corneal infection. The polypeptides can also be used to aid wound healing and epithelial cell proliferation, to prevent skin aging due to sunburn, to maintain organs before transplantation, for supporting cell culture of primary tissues, to regenerate tissues and in chemotaxis. The polypeptides can also be used as a food additive or preservative to increase or decrease storage capabilities, fat content, lipid, protein, carbohydrate, vitamins, minerals, co-factors and other nutritional components. AAF22307 to AAF22315 and AAB3048 represent sequences used in the exemplification of the present invention

| | | | |
|----|-----|---|-----|
| Qy | 508 | CRAAAGATTACTGACACTTGGATTCCATTAAAGAAAAGCTTAACCAACTCGGATTCACA | 567 |
| Db | 283 | CAAAAGATTACTGACACTTGGATTCCATTAAAGAAAAGCTTAACCAACTCGGATTCACA | 342 |
| Qy | 568 | AATTTTGACCTATGTCCTGCTTCCACCATCCTGGACTACTGGACATACCTGGTTCT | 627 |
| Db | 343 | AATTTTGACCTATGTCCTGCTTCCACCATCCTGGACTACTGGACATACCTGGTTCT | 402 |
| Qy | 628 | CTTACAGTCCACCTCTTCTTGAGGTGTACATGGATTGTTTAAGCAAACCTATAAAC | 687 |
| Db | 403 | CTTACAGTCCACCTCTTCTTGAGGTGTACATGGATTGTTTAAGCAAACCTATAAAC | 462 |
| Qy | 688 | ATCAGCTCTAACAGTGGCCAATTTCGGAGTCCTGTGCAACGGAGGGTGAACCA | 747 |
| Db | 463 | ATCAGCTCTAACAGTGGCCAATTTCGGAGTCCTGTGCAACGGAGGGTGAACCA | 522 |
| Qy | 748 | GCACCTTCTGATAGACTCTCTACTCTGTCACCCAGGTGGAGGCCAG | 797 |
| Db | 523 | GCACCTTCTGATAGACTCTCTACTCTGTCACCCAGGTGGAGGCCAG | 572 |

8

¹⁴C-68431 standard; cDNA; 462 BP.

XX AAS68431;
 XX AC
 XX DT 13-FEB-2002 (first entry)
 XX DB encoding novel human diagnostic protein #4235.
 DB Human; chromosome mapping; gene mapping; gene therapy; forensics;
 XX KW

| | | | |
|----|--|----------|--|
| KW | food supplement; medical imaging; diagnostic; genetic disorder; ss. | Db | 206 AATCAGGTTCTGGGGGTGTCCTCACTGGAAAGCTACAGGTACGGCAGGTTAACCTTC 147 |
| XX | Homo sapiens. | Qy | 323 ACTGGGGCCGCTGATGACCACCGCTGGAGCATAGTAGGAGTGAGCTATGCCTG 382 |
| XX | WO200175067-A2. | Db | 146 ACTGGGGTCCGCTGATGACCACCGCTGGAGCATAGTAGGAGTGAGCTATGCCTG 87 |
| XX | 11-OCT-2001. | Qy | 383 CAGAGCTCCATTGTTCACTGGAAATTGAGACAATACCCAGTTGTTGAGGCCAGTC 442 |
| PD | 30-MAR-2001; 2001WO-US008631. | Db | 86 CAGAGCTCCATTGTTCACTGGAAATTGAGACAATACCCAGTTGTTGAGGCCAGTC 27 |
| XX | 31-MAR-2000; 2000US-00540217. | Qy | 443 ATGAAACCAGATGGACTGGCTGTCTG 468 |
| PR | 23-AUG-2000; 2000US-00649167. | Db | 26 ATGAAACCAGATGGACTGGCTGTCTG 1 |
| XX | (HYSEQ -) HYSEQ INC. | | |
| XX | PI Dermanac RT, Liu C, Tang YT; | RESULT 9 | |
| XX | WPI: 2001-639362/73. | ACH26864 | |
| DR | P-PSDB; ABG04244. | ID | ACH26864 standard; cDNA; 496 BP. |
| XX | | XX | |
| PT | New isolated polynucleotide and encoded polypeptides, useful in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits and to assess biodiversity. | AC | ACH26864; |
| PT | | XX | DT 13-OCT-2003 (first entry) |
| PT | | XX | DE Human adult ovary cDNA #5244. |
| XX | | XX | KW Human; ss; sequencing by hybridisation; SBH; expressed sequence tag; EST; genome mapping; biodiversity; genetic disorder. |
| PS | Claim 1; SEQ ID NO 4235; 103pp; English. | OS | Homo sapiens. |
| XX | The invention relates to isolated polynucleotide (I) and polypeptide (II) sequences. (I) is useful as hybridisation probes, polymerase chain reaction (PCR) primers, oligomers, and for chromosome and gene mapping, and in recombinant production of (II). The polynucleotides are also used in diagnostics as expressed sequence tags for identifying expressed genes. (I) is useful in gene therapy techniques to restore normal activity of (II) or to treat disease states involving (II). (II) is useful for generating antibodies against it, detecting or quantitating a polypeptide in tissue, as molecular weight markers and as a food supplement. (II) and its binding partners are useful in medical imaging of sites expressing (II). (I) and (II) are useful for treating disorders involving aberrant protein expression or biological activity. The polypeptide and polynucleotide sequences have applications in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits to assess biodiversity and to produce other types of data and products dependent on DNA and amino acid sequences. AAS6197-AAS9564 represent novel human diagnostic coding sequences of the invention. Note: The sequence data for this patent did not appear in the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences | PN | US2003073623-A1. |
| XX | XX | XX | US |
| XX | XX | XX | US |
| XX | XX | XX | 2003-APR-07. |
| XX | XX | XX | 30-JUL-2001; 2001US-00918995. |
| XX | XX | XX | 30-JUL-2001; 2001US-00918995. |
| XX | XX | XX | XX |
| XX | XX | XX | (DRMANAC R T. |
| XX | XX | XX | PA (LABA/) |
| XX | XX | XX | LABAT I. |
| XX | XX | XX | PA (STAC/) |
| XX | XX | XX | STACHE-CRAIN B. |
| XX | XX | XX | PA (DICK/) |
| XX | XX | XX | DICKSON M C. |
| XX | XX | XX | PA (JONE/) |
| XX | XX | XX | JONES L W. |
| XX | XX | XX | PA |
| XX | XX | XX | PI DRMANAC RT, Labat I, Stache-Crain B, Dickson MC, Jones LW; |
| XX | XX | XX | DR WPI; 2003-615964/58. |
| SQ | Sequence 462 BP; 113 A; 115 C; 121 T; 0 U; 0 Other; | PT | New polynucleotide sequences obtained from various cDNA libraries, useful as hybridization probes, as oligomers for PCR, for chromosome and gene mapping, in the recombinant production of protein, or in generating antisense DNA or RNA. |
| XX | Best Local Similarity 99.8%; Pred. No. 4.6e-134; Matches 445; Conservative 0; Mismatches 1; Indels 0; Gaps 0; | PT | Claim 1; SEQ ID NO 14076; 44pp; English. |
| XX | XX | XX | XX |
| Qy | 23 CCCGAGGAATGTCAGGCTCAGCTGGATAACGGACACGGCTCATCTACT 82 | CC | The invention relates to an isolated polynucleotide comprising any one of 3804 CDNA sequences, appearing as ACH12789-ACH50831, whose sequence was determined by the technique of SBH (sequencing by hybridisation). Also included is a purified polypeptide comprising a sequence corresponding to a reading frame of the novel polynucleotide. The nucleic acid sequences are useful in diagnostics as expressed sequence tags (EST) for identifying expressed genes or for physical mapping of the human genome, in forensics, in assessing biodiversity and other traits. The nucleotide sequences are also useful as hybridisation probes, as oligomers for PCR, for chromosome and gene mapping, in the recombinant production of protein, or in generating antisense DNA or RNA. The purified polypeptide is useful for generating antibodies specific for it. The present sequence is one of the 3803 isolated cDNA/EST sequences. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from USPTO at seqdata.uspto.gov/sequence.html?docID=20030073623 |
| Db | 446 CCCGAGGAATGTCAGGCTCAGCTGGATAACGGACACGGCTCATCTACT 387 | CC | XX |
| Qy | 83 GGAGGAAATTTCCTATGCTGATGGTCAATTGAGATAAACCA 142 | CC | XX |
| Db | 386 GGAAAGGAAATTTCCTATGCTGATGGTCAATTGAGATAAACCA 327 | CC | XX |
| Qy | 143 AAGAAATGAAATTGACTCTCCCTCGACCACTAGTAAAGTACGGCAACTCTAG 202 | CC | XX |
| Db | 326 AAGAAATGAAATTGACTCTCCCTCGACCACTAGTAAAGTACGGCAACTCTAG 267 | CC | XX |
| Qy | 203 CTAAATTCATGCAACAGCGCCATTCTCAATTGATGACAGAGAACAA 262 | CC | XX |
| Db | 266 CTAAAATCATGCAACAGCGCCATTCTCAATTGATGACAGAGAACAA 207 | CC | XX |
| Qy | 263 AATCAGTTCTGGTGGTCTCACTGGAAAGCTACGGCAGGTTAACCTTC 322 | CC | XX |

| ISQ | Sequence | 496 BP; 128 A; 122 C; 121 G; 119 T; 0 U; 6 Other; |
|-----------|--|---|
| | Query Match | 48.8%; Score 44.4; DB 8; Length 496; |
| | Best Local Similarity | 99.8%; Pred. No. 4.8e-134; |
| | Matches 445; Conservative 0; Mismatches 1; Indels 0; Gaps 0; | |
| QY | 23 CCCGAGGGACCATGTGAGGCTCACTGGGGATACCGGAGGCACAGGTCTTATTACT 82 | |
| | 51 CCCGAGGGACCATGTGAGGCTCACTGGGGATACCGGAGGCACAGGTCTTATTACT 110 | |
| QY | 83 GGAAAGGAATTTCCTTCCATTGCTGATGCCAATTCCTGATGGTCAAGTTGAGATAAACCA 142 | |
| Db | 111 GGAAAGGAATTTCCTTCCATTGCTGATGCCAATTCCTGATGGTCAAGTTGAGATAAACCA 170 | |
| QY | 143 AAGAACTGAAATATGACTCTCCCTGACCACTTAGTATCAAGTAAGCTGACCCAAAGCTCAG 202 | |
| Db | 171 AAGAACTGAAATATGACTCTCCCTGACCACTTAGTATCAAGTAAGCTGACCCAAAGCTCAG 230 | |
| QY | 203 CTAAAATCATGCAAGCGCCATTCCCTTCAATGTTGACTTGATGACAGAGAAC 262 | |
| Db | 231 CTAAAATCATGCAAGCGCCATTCCCTTCAATGTTGACTTGATGACAGAGAAC 290 | |
| QY | 263 AATCAGTTCTCGTGTGGCTCTCACTGGAAAGCTACAGGTACGGCTTACACTTC 322 | |
| Db | 291 AATCAGTTCTCGTGTGGCTCTCACTGGAAAGCTACAGGTACGGCTTACACTTC 350 | |
| QY | 323 ACTGGGGTCCGTGTGATGACCAACGGCTCGGCACTAGTAGATGGTGTGAGCTATGCTG 382 | |
| Db | 351 ACTGGGGTCCGTGTGATGACCAACGGCTCGGCACTAGTAGATGGTGTGAGCTATGCTG 410 | |
| QY | 383 CAGAGCTCCATGTTGTCACTGGAAATTCAAGAAATAACCCCCAGGTTGTGAGGAAGCTC 442 | |
| Db | 411 CAGAGCTCCATGTTGTCACTGGAAATTCAAGAAATAACCCCCAGGTTGTGAGGAAGCTC 470 | |
| QY | 443 ATGAAACCAGATGGACTGGCTGCTCTG 468 | |
| Db | 471 ATGAAACCAGATGGACTGGCTGCTCTG 496 | |
| RESULT 10 | | |
| XX | AAI92460 standard; cDNA; 462 BP. | |
| XX | AAI92460; | |
| AC | | |
| XX | 06-NOV-2001 (first entry) | |
| DT | | |
| XX | | |
| DE | Human polynucleotide SEQ ID NO 12520. | |
| XX | Human; cytokine; cell proliferation; cell differentiation; gene therapy; vaccine; peptide therapy; stem cell growth factor; hematopoiesis; tissue growth factor; immunomodulatory; cancer; leukaemia; nervous system disorders; arthritis; inflammation; ss. | |
| KW | | |
| XX | Homo sapiens. | |
| OS | | |
| XX | | |
| PN | W0200164935-A2. | |
| XX | 07-SEP-2001. | |
| PD | | |
| XX | 26-FEB-2001; 2001WO-US004927. | |
| XX | 28-FEB-2000; 2000US-00515126. | |
| PR | 18-MAY-2000; 2000US-00577409. | |
| XX | | |
| PA | (HYSE-) HYSEQ INC. | |
| XX | | |
| PI | Tang YT, Liu C, Drmanac RT; | |
| XX | | |
| DR | WPI; 2001-514838/56. | |
| DR | P-PSDB; AAO12329. | |
| XX | Isolated nucleic acids and polypeptides, useful for preventing diagnosing | |
| PT | | |

| Result No. | Score | Query | Match | Length | DB ID | Description |
|------------|-------|-------|-------|--------|--------|--------------------|
| 1 | 804 | 62.4 | 261 | 1 | CRH1 | carbonate dehydrat |
| 2 | 804 | 62.4 | 261 | 2 | JN0836 | carbonate dehydrat |
| 3 | 800 | 62.1 | 260 | 1 | CRH01D | carbonate dehydrat |
| 4 | 800 | 62.1 | 261 | 2 | JN0835 | carbonate dehydrat |
| 5 | 788.5 | 61.2 | 259 | 1 | CRSH2 | carbonate dehydrat |
| 6 | 788.5 | 61.2 | 260 | 1 | CRH02 | carbonate dehydrat |
| 7 | 785.5 | 60.9 | 259 | 1 | CRB02 | carbonate dehydrat |
| 8 | 783.5 | 60.8 | 260 | 2 | T08463 | carbonate dehydrat |
| 9 | 780 | 60.5 | 260 | 1 | CRMQ1R | carbonate dehydrat |
| 10 | 778.5 | 60.4 | 260 | 2 | JG2580 | carbonate dehydrat |
| 11 | 761.5 | 59.1 | 259 | 1 | CRRB2 | carbonate dehydrat |
| 12 | 754.5 | 58.5 | 259 | 2 | A22612 | carbonate dehydrat |
| 13 | 754 | 58.5 | 259 | 1 | A23344 | carbonate dehydrat |
| 14 | 749.5 | 58.1 | 261 | 1 | CRMS2 | carbonate dehydrat |
| 15 | 748.5 | 58.1 | 260 | 2 | JH0527 | carbonate dehydrat |
| 16 | 748.5 | 58.1 | 260 | 2 | A33641 | carbonate dehydrat |
| 17 | 747.5 | 58.0 | 260 | 1 | CRH03 | carbonate dehydrat |
| 18 | 747.5 | 58.0 | 260 | 2 | A15551 | carbonate dehydrat |
| 19 | 717.5 | 55.7 | 235 | 2 | A22962 | carbonate dehydrat |
| 20 | 689 | 53.5 | 260 | 1 | CRH07 | carbonate dehydrat |
| 21 | 589 | 45.7 | 305 | 1 | CRH05 | carbonate dehydrat |
| 22 | 581.5 | 45.1 | 304 | 2 | J59261 | carbonate dehydrat |
| 23 | 549.5 | 42.6 | 298 | 2 | S12579 | carbonate dehydrat |
| 24 | 453 | 35.5 | 200 | 2 | A60519 | carbonate dehydrat |
| 25 | 453 | 35.1 | 303 | 2 | S12867 | carbonate dehydrat |
| 26 | 451 | 35.0 | 290 | 2 | JN0576 | cell surface-bindi |
| 27 | 392 | 30.4 | 304 | 1 | CRVZ7P | cell surface-bindi |
| 28 | 386 | 29.9 | 304 | 2 | CRVZW | cell surface-bindi |
| 29 | 386 | 29.9 | 304 | 2 | H72162 | F8L protein - vari |

| RESULT 1 | | | | | | | | | |
|---|--|--|--|--|--|--|--|--|--|
| CRH1 | | | | | | | | | |
| carbonate dehydratase (EC 4.2.1.1) I [validated] - human | | | | | | | | | |
| N;Alternate names: carbonic anhydrase I | | | | | | | | | |
| C;Species: Homo sapiens (man) | | | | | | | | | |
| C;Date: 07-May-1981 #Sequence revision 05-May-1995 #Text change 08-Dec-2000 | | | | | | | | | |
| C;Accession: JQ0786; A26573; R90668; A91180; A92128; B92147; A01138 | | | | | | | | | |
| R;Ilowe, N.; Brady, H.J.M.; Barlow, J.H.; Sowden, J.C.; Edwards, M.; Butterworth, P.H.W. | | | | | | | | | |
| Gene 93, 277-283, 1990 | | | | | | | | | |
| A;Title: Structure and methylation patterns of the gene encoding human carbonic anhydras | | | | | | | | | |
| A;Reference number: JQ0786; MUID:3103039; PMID:2121614 | | | | | | | | | |
| A;Accession: JQ0786 | | | | | | | | | |
| A;Molecule type: DNA | | | | | | | | | |
| A;Residues: 1-261 <LOW> | | | | | | | | | |
| A;Cross-references: GB:M33987; NID:9179792; PID:MA51910-1; PMID:917993 | | | | | | | | | |
| A;Experimental source: erythrocyte | | | | | | | | | |
| R;Barlow, J.H.; Ilowe, N.; Edwards, Y.H.; Butterworth, P.H.W. | | | | | | | | | |
| Nucleic Acids Res. 15, 2386, 1987 | | | | | | | | | |
| A;Title: Human carbonic anhydrase I cDNA. | | | | | | | | | |
| A;Reference number: A26573; PMID:2121614 | | | | | | | | | |
| A;Molecule type: mRNA | | | | | | | | | |
| A;Residues: 1-261 <BAR> | | | | | | | | | |
| A;Cross-references: GB:AO5014; NID:92959; PID:CA82663-1; PMID:929606 | | | | | | | | | |
| A;Note: the authors translated the codon GAG for residue 118 as GLY | | | | | | | | | |
| R;Giraud, N.; Marrig, C.; Laurent-Tabusse, G. | | | | | | | | | |
| Biochimie 56, 1043, 1974 | | | | | | | | | |
| A;Title: Structure primaire de l'anhydrase carbonique erythrocytaire B humaine. III. Seq | | | | | | | | | |
| A;Reference number: A90668; MUID:7509T08; PMID:2171796 | | | | | | | | | |
| A;Accession: A50668 | | | | | | | | | |
| A;Molecule type: protein | | | | | | | | | |
| A;Residues: 2-74, PRO, 77-261 <GIR> | | | | | | | | | |
| R;Andersson, B.; Nyman, P.O.; Strid, L. | | | | | | | | | |
| Biochem. Biophys. Res. Commun. 48, 670-677, 1972 | | | | | | | | | |
| A;Title: Amino acid sequence of human erythrocyte carbonic anhydrase B. | | | | | | | | | |
| A;Reference number: A90180; MUID:72243008; PMID:4625866 | | | | | | | | | |
| A;Accession: A90180 | | | | | | | | | |
| A;Molecule type: protein | | | | | | | | | |
| A;Residues: 20-74, ND, 77-261 <AND> | | | | | | | | | |
| R;Lin, K.T.D.; Deutsch, H.F. | | | | | | | | | |
| J. Biol. Chem. 248, 1885-1893, 1973 | | | | | | | | | |
| A;Title: Human carbonic anhydrases. XI. The complete primary structure of carbonic anhydr | | | | | | | | | |
| A;Reference number: A92128 | | | | | | | | | |
| A;Molecule type: protein | | | | | | | | | |
| A;Residues: 12-26, DQN, 30-165, 'E', 167-261 <LIN> | | | | | | | | | |
| A;Note: this sequence has been revised in reference A92147 | | | | | | | | | |
| R;Lin, K.T.D.; Deutsch, H.F. | | | | | | | | | |
| J. Biol. Chem. 249, 2329-2337, 1974 | | | | | | | | | |
| A;Title: Human carbonic anhydrases. XII. The complete primary structure of the C isozyme | | | | | | | | | |
| A;Reference number: A92147; MUID:74143468; PMID:4207120 | | | | | | | | | |
| A;Accession: B92147 | | | | | | | | | |

ALIGNMENTS

Searched:

Total number of hits satisfying chosen parameters:

283366

Run on: September 9, 2004, 15:06:59 ; Search time 40 Seconds
 (without alignments)
 (5.1959 Million cell updates/sec)

Title: US-10-069-434-1
 Perfect score: 1289
 Sequence: 1 MSRLSWGYREHNGPIHWKEF.....QLAKFRSLCTAEGEAAFL 242

Scoring table: BLOSUM62
 Gapext 0.5 , Gapext 0.5

Searched:

A;Molecule type: protein-2nd
A;Residues: 2-74, ND-'77-261 >
R;Kannan, K.K.; Notstrand, B.; Friisborg, K.; Lovgren, S.; Ohlsson, A.; Petersen, M.
Proc. Natl. Acad. Sci. U.S.A. 72, 51-55, 1975
A;Title: Crystal structure of human erythrocyte carbonic anhydrase B. Three-dimensional
A;Reference number: A93803; PMID:75120422; PMID:804171
A;Annotation: X-ray crystallography, 2.2 angstroms
C;Genetics:
A;Gene: GNB1CA1
A;Cross-references: GDB:119047; OMIM:114800
A;Map location: 8q13.8;q22.1
A;Introns: 13/1; 79/1; 118/3; 150/3; 171/3; 223/3
A;Note: the first intron occurs before the initiator codon
C;Function:
A;Description: catalyzes the reversible dissociation of carbonic acid to carbon dioxide
A;Note: This form is predominantly expressed in erythrocytes
C;Superfamily: carbonate dehydratase; carbonic anhydrase homology
C;Keywords: acetylated amino end; carbon-oxygen lyase; hydro-lyase; metalloprotein; zinc
P;6-261/Product: carbonate dehydratase I #status experimental
P;6-261/Domain: carbonic anhydrase homology <CAH>
P;7/2/Modified site: acetylated amino end (Ala) (in mature form) #status experimental
P;95/97,120/Binding site: zinc (His) #status experimental

Query Match Score 804; DB 1; Length 261;
Best Local Similarity 59.4%; Pred. No. 1..3e-60;
Matches 142; Conservative 44; Mismatches 53; Indels 0; Gaps 0;

Qy 1 MSRLSWGTYREHNGPIHWKEFFPIADGDOQSPIBIKTKEVYKDSSLRPLPSIKYDPSAKII 60
Db 1 MASPDWGTDKNGEQWMSKLYPANGNNSQSPVDTKTSKHDTSLKPSVSYNPATAKEI 60

Qy 61 SNSGHSPNVDDEDDTENKSYLVRGSPLTGSYRLRQVHLWGSADDHGSEHIVDGYSAAELH 120
Db 61 INYGHSPNVDDEDDTENKSYLVRGSPLTGSYRLRQVHLWGSADDHGSEHIVDGYSAAELH 120

Qy 121 VTHWNSDKYPSFVEAAHEPDGLAVLGFLQIGEPNSQOKITDLSIKEKGKOTRFNF 180
Db 121 VAHKNSAKYSSLAEEASKADGLAVLGVLKVGEANPKLQVLDALQAKTKGRAPFTNF 180

Qy 181 DLSLILPPSWDYWTYPGSLTVPPPLESTWIVLKOPINISSQOLAKFSLCTAEGEA 239
Db 181 DPSTLILPSSLDFWTYPGSLTHPPLESTWIVKICKESISVSSEQLAQFSSLNSVEGDNA 239

RESULT 2
JN0836
carbonate dehydratase (EC 4.2.1.1) I - gorilla
N;Alternative names: carbonic anhydrase
C;Species: Gorilla (gorilla)
C;Date: 03-May-1994 #sequence_revision 03-May-1994 #text_change 07-May-1999
C;Accession: JN0836
R;Epperly, B.R.; Bergenhem, N.C.H.; Venta, P.J.; Tashian, R.E.
Gene 131, 249-253, 1993
A;Title: Characterization of the genes encoding carbonic anhydrase I of chimpanzee and gorilla
A;Reference number: JN0835; PMID:94010316; PMID:8406018
A;Accession: JN0836
A;Molecule type: mRNA
A;Residues: 1-261 <EP>>
A;Cross-references: GB:L11622
A;Note: the authors translated codon GAG for residue 206 as Leu
C;Genetics:
A;Introns: 12/1; 78/1; 117/3; 148/3; 170/3; 222/3
C;Superfamily: carbonate dehydratase; carbonic anhydrase homology
C;Keywords: carbon-oxygen lyase; hydro-lyase
P;6-261/Domain: carbonic anhydrase homology <CAH>

Query Match Score 804; DB 2; Length 261;
Best Local Similarity 59.8%; Pred. No. 1..3e-60;
Matches 143; Conservative 43; Mismatches 53; Indels 0; Gaps 0;

Qy 1 MSRLSWGTYREHNGPIHWKEFFPIADGDOQSPIBIKTKEVYKDSSLRPLPSIKYDPSAKII 60
Db 1 MASPDWGDDKNGPEQWMSKLYPANGNNSQSPVDTKTSKHDTSLKPSVSYNPATAKEI 60

Qy 61 SNSGHSPNVDDEDDTENKSYLVRGSPLTGSYRLRQVHLWGSADDHGSEHIVDGYSAAELH 120
Db 61 INYGHSPNVDDEDDTENKSYLVRGSPLTGSYRLRQVHLWGSADDHGSEHIVDGYSAAELH 120

Qy 121 VTHWNSDKYPSFVEAAHEPDGLAVLGFLQIGEPNSQOKITDLSIKEKGKOTRFNF 180
Db 121 VAHKNSAKYSSLAEEASKADGLAVLGVLKVGEANPKLQVLDALQAKTKGRAPFTNF 180

RESULT 3
CRHOID
carbonate dehydratase (EC 4.2.1.1) I - horse
N;Alternative names: carbonic anhydrase I
C;Species: Equus caballus (domestic horse)
C;Date: 22-May-1981 #sequence_revision 22-May-1981 #text_change 11-Nov-1996
C;Accession: A01140
R;Jabusch, J.R.; Bray, R.P.; Deutsch, H.F.
J. Biol. Chem. 255, 9196-9204, 1980
A;Title: Sequence of the low activity equine erythrocyte carbonic anhydrase and delineat
A;Reference number: A01140; MUID:81006999; PMID:6739361
A;Contents: D-isozyme
A;Accession: A01140
A;Molecule type: Protein
A;Residues: 1-260 <JAB>
A;Note: 65-Gly, 115-His, 157-Gly, 212-Tyr, and 224-Ala were also found in the electropho
1 isozyme has 183-Arg; and the B isozyme has 183-Arg and 222-Arg
C;Superfamily: carbonic dehydratase; carbonic anhydrase homology
C;Keywords: carbon-oxygen lyase; hydro-lyase; zinc
P;5-260/Domain: carbonic anhydrase homology <CAH>
F;94, 96,119/Binding site: zinc (His) #status predicted

Query Match Score 800; DB 0; Length 260;
Best Local Similarity 62.0%; Pred. No. 2..8e-60;
Matches 145; Conservative 35; Mismatches 54; Indels 0; Gaps 0;

Qy 6 WGYREHNGPIHWKEFFPIADGDOQSPIBIKTKEVYKDSSLRPLPSIKYDPSAKII 65
Db 5 WGDSPBGPZEWKLYPLABGBQSPIDKTSKHDTSLKPSVSYDPATAKEIVANGH 64

Qy 66 SFNYDFDDTENKSYLVRGSPLTGSYRLRQVHLWGSADDHGSEHIVDGYSAAELHVVHN 125
Db 65 SFQKFEDSDNRSTLKDGLPLPSYRQLQFHFGSTDYDGETTVDGKYSAELHLVHN 124

Qy 126 SDKYPSFVEAAHEPDGLAVLGFLQIGEPNSQOKITDLSIKEKGKOTRFNFDSLISL 185
Db 125 SSKSFSDFAASSQDAGLTLGVLMKVGEANPKLQVLDALNEYTKGKAKPFKNFDPSLL 184

Qy 186 LPPSWDYWTYPGSLTVPPPLESTWIVLKOPINISSQOLAKFSLCTAEGEA 239
Db 185 LPSPDWTYPSGSLTHPPLESTWIVKICKESISVSSEQLAQFSSLNSVEGKA 238

RESULT 4
JN0335
carbonate dehydratase (EC 4.2.1.1) I - chimpanzee
N;Alternative names: carbonic anhydrase
C;Species: Pan troglodytes (chimpanzee)
C;Date: 03-May-1994 #sequence_revision 03-May-1994 #text_change 07-May-1999
C;Accession: JN0335
R;Epperly, B.R.; Bergenhem, N.C.H.; Venta, P.J.; Tashian, R.E.
Gene 131, 249-253, 1993
A;Title: Characterization of the genes encoding carbonic anhydrase I of chimpanzee and gorilla
A;Reference number: JN0335; MUID:94010316; PMID:8406018
A;Accession: JN0335
A;Molecule type: mRNA
A;Residues: 1-261 <EP>>
A;Cross-references: GB:L11621
A;Note: the authors translated codon GRG for residue 206 as Leu
C;Cross-references: GB:L11621 <EP>>

C;Genetics:
A;Introns: 12/1; 78/1; 117/3; 148/3; 170/3; 222/3
C;Superfamily: carbonic anhydrase; carbonic anhydrase homology
C;Keywords: carbon-oxygen lyase; hydro-lyase
F;6-267/Domain: carbonic anhydrase homology <CAH>

Query Match 62.1%; Score 800; DB 2; Length 261;

Best Local Similarity 58.6%; Pred. No. 2.8e-60;
Matches 140; Conservative 46; Mismatches 53; Indels 0; Gaps 0;

Db 1 MSRLSWGYREHNGPIHWKEFFPIADGDOQSPLEIKTKVKYDSSRPLSIKYDPPSSAKI 60

Db 1 MSAPDWGQDDPQEWPQWANGKPSLTPSYNTPATAKE 60

Qy 61 SNSGHFSNVDFTDTENKSVERGPGLTGSYRLRQHHLGASDHGSEITVGDVSYAAELH 120

Db 61 INYGHFSFHVNFNEUDNRSVLYKGPFPSDSYRLQFHEFWGSTNEHGSEHTVGKVSYAAELH 120

Qy 121 VVHNNSDKXPSFVEAAHEDPGIAVLGVFLQIGEPNSOLQKQTDTLDSTKEKGKQTRETFNF 180

Db 121 IAHWSNAYSNLDEEAKADGLAVIGVMKGVDANPQLQKIRTKRKPPTNF 180

Qy 181 DLSLLPPSDWDYTYPGSLTVPPLESVTWILKOPINISSQQLAKPFRSLCTAEGFAA 239

Db 181 DPSTLPPSDWFYTYPGSLTHPPLESVTWICKESSISQQLAQFRSLLSNVEGDNA 239

Query Match 62.1%; Score 800; DB 2; Length 261;

Best Local Similarity 58.6%; Pred. No. 2.8e-60;
Matches 140; Conservative 46; Mismatches 53; Indels 0; Gaps 0;

Db 1 MSRLSWGYREHNGPIHWKEFFPIADGDOQSPLEIKTKVKYDSSRPLSIKYDPPSSAKI 60

Db 1 MSAPDWGQDDPQEWPQWANGKPSLTPSYNTPATAKE 60

Qy 61 SNSGHFSNVDFTDTENKSVERGPGLTGSYRLRQHHLGASDHGSEITVGDVSYAAELH 120

Db 61 INYGHFSFHVNFNEUDNRSVLYKGPFPSDSYRLQFHEFWGSTNEHGSEHTVGKVSYAAELH 120

Qy 121 VVHNNSDKXPSFVEAAHEDPGIAVLGVFLQIGEPNSOLQKQTDTLDSTKEKGKQTRETFNF 180

Db 121 IAHWSNAYSNLDEEAKADGLAVIGVMKGVDANPQLQKIRTKRKPPTNF 180

Qy 181 DLSLLPPSDWDYTYPGSLTVPPLESVTWILKOPINISSQQLAKPFRSLCTAEGFAA 239

Db 181 DPSTLPPSDWFYTYPGSLTHPPLESVTWICKESSISQQLAQFRSLLSNVEGDNA 239

Query Match 62.1%; Score 800; DB 2; Length 261;

Best Local Similarity 58.6%; Pred. No. 2.8e-60;
Matches 140; Conservative 46; Mismatches 53; Indels 0; Gaps 0;

Db 1 MSRLSWGYREHNGPIHWKEFFPIADGDOQSPLEIKTKVKYDSSRPLSIKYDPPSSAKI 60

Db 1 MSAPDWGQDDPQEWPQWANGKPSLTPSYNTPATAKE 60

Qy 61 SNSGHFSNVDFTDTENKSVERGPGLTGSYRLRQHHLGASDHGSEITVGDVSYAAELH 120

Db 61 INYGHFSFHVNFNEUDNRSVLYKGPFPSDSYRLQFHEFWGSTNEHGSEHTVGKVSYAAELH 120

Qy 121 VVHNNSDKXPSFVEAAHEDPGIAVLGVFLQIGEPNSOLQKQTDTLDSTKEKGKQTRETFNF 180

Db 121 IAHWSNAYSNLDEEAKADGLAVIGVMKGVDANPQLQKIRTKRKPPTNF 180

Qy 181 DLSLLPPSDWDYTYPGSLTVPPLESVTWILKOPINISSQQLAKPFRSLCTAEGFAA 239

Db 181 DPSTLPPSDWFYTYPGSLTHPPLESVTWICKESSISQQLAQFRSLLSNVEGDNA 239

Query Match 62.1%; Score 800; DB 2; Length 261;

Best Local Similarity 58.6%; Pred. No. 2.8e-60;
Matches 140; Conservative 46; Mismatches 53; Indels 0; Gaps 0;

Db 1 MSRLSWGYREHNGPIHWKEFFPIADGDOQSPLEIKTKVKYDSSRPLSIKYDPPSSAKI 60

Db 1 MSAPDWGQDDPQEWPQWANGKPSLTPSYNTPATAKE 60

Qy 61 SNSGHFSNVDFTDTENKSVERGPGLTGSYRLRQHHLGASDHGSEITVGDVSYAAELH 120

Db 61 INYGHFSFHVNFNEUDNRSVLYKGPFPSDSYRLQFHEFWGSTNEHGSEHTVGKVSYAAELH 120

Qy 121 VVHNNSDKXPSFVEAAHEDPGIAVLGVFLQIGEPNSOLQKQTDTLDSTKEKGKQTRETFNF 180

Db 121 IAHWSNAYSNLDEEAKADGLAVIGVMKGVDANPQLQKIRTKRKPPTNF 180

Qy 181 DLSLLPPSDWDYTYPGSLTVPPLESVTWILKOPINISSQQLAKPFRSLCTAEGFAA 239

Db 181 DPSTLPPSDWFYTYPGSLTHPPLESVTWICKESSISQQLAQFRSLLSNVEGDNA 239

RESULT 6
CRH02 carbonate dehydratase (EC 4.2.1.1) II [validated] - human

N;Alternate names: carbonic anhydrase II; hepatic carbonic anhydrase
C;Species: Homo sapiens (man)

C;Date: 07-May-1981 #sequence revision 05-May-1995 #text change 15-Sep-2000
C;Accession: A27175; A23202; A92147; I37214; I51871; A01141

C;Title: Cloning, expression, and sequence homologies of cDNA for human carbonic anhydrase

A;Reference number: A27175; MUID:80885190; PMID:3121496

A;Accession: A27175
A;Molecule type: mRNA

A;Residues: 1-260 <MR>

A;Cross-references: GB:J03037; NID:9179771; PID:AAA51908.1; PMID:9179772

R;Ventz, P.J.; Montgomery, C.; Hewett-Emmett, D.; Tashian, R.E.

Biochim. Biophys. Acta 826, 195-201, 1985

A;Title: Comparison of the 5' regions of human and mouse carbonic anhydrase II genes and

A;Reference number: A90655; MUID:8607780; PMID:300449

A;Accession: A27175
A;Molecule type: DNA

A;Residues: 1-77 <VN>

A;Cross-references: GB:X03251; GB:M18100; PMID:9179778; PID:AAA51909.1; PMID:9179778

R;Henderson, L.E.; Henrikson, D.; Nyman, P.O.

J. Biol. Chem. 251, 5457-5463, 1976

A;Title: The primary structure of human carbonic anhydrase C.

A;Reference number: A92194; MUID:77006079; PMID:823150

A;Accession: A92194
A;Molecule type: protein

A;Residues: 2-260 <HN>

A;Cross-references: EMBL:2-260 <HN>

R;Lin, K.T.D.; Deutscher, H.F.

J. Biol. Chem. 249, 2329-2337, 1974

A;Title: Human carbonic anhydrases. XII. The complete primary structure of the C isozyme

A;Reference number: A92147; MUID:74143468; PMID:4207120

A;Status: translated from GB/EMBL/DDJB

A;Molecule type: mRNA

A;Residues: 1-260 <R33>

A;Cross-references: EMBL:Y00339; NID:929586; PID:CAA68426.1; PMID:929587

A;Experimental source: liver

A;Note: submitted to the EMBL/GenBank/DBJ databases by David Hewett-Emmett 01-JUL-1987

R;Ventz, P.J.; Welty, R.J.; Johnson, T.M.; Sily, W.S.; Tashian, R.E.

Am. J. Hum. Genet. 49, 1082-1090, 1991

A;Title: Carbonic anhydrase II deficiency syndrome in a Belgian family is caused by a p.o.e.

A;Reference number: I51863; MUID:9026087; PMID:1928091

A;Accession: I51863
A;Status: translation not shown; translated from GB/EMBL/DDJB

A;Molecule type: DNA

A;Residues: 1-260 <ES3>

A;Cross-references: GB:M77181; NID:9179778; PID:AAA51909.1; PMID:9179780

A;Note: the complete nucleotide sequence is not shown

R;Hu, P.Y.; Ernst, A.R.; Sily, W.S.; Ventz, P.J.; Skaggs, L.A.; Tashian, R.E.

Am. J. Hum. Genet. 54, 602-608, 1994

A;Title: Carbonic anhydrase II deficiency: single base deletion in exon 7 is the predomi

A;Reference number: I51871; MUID:94175074; PMID:8128957

A;Accession: 151871
A;Status: translated from GB/EMBL/DDJB

A;Molecule type: mRNA

A;Residues: 225-242 <RE2>

A;Cross-references: GB:S8526; NID:954550; PID:NAB30170.1; PMID:954551

A;Note: wild type shown; mutant contains frameshift after residue 226

R;Ericsson, A.B.; Jones, T.A.; Lijas, A.

submitted to the Brookhaven Protein Data Bank, February 1989

A;Reference number: A50085; PDB:1C2A

RESULT 5
CRH02 carbonate dehydratase (EC 4.2.1.1) II - sheep (tentative sequence)

N;Alternate names: carbonic anhydrase II

C;Species: Ovis orientalis aries (domestic sheep)

C;Date: 31-Mar-1981 #sequence revision 31-Mar-1981 #text_change 31-Mar-2000

C;Accession: A01145
R;Tanis, R.J.; Ferrell, R.E.

Biochim. Biophys. Acta 371, 534-549, 1974

A;Title: Amino acid sequence of sheep carbonic anhydrase C.

A;Reference number: A90598; MUID:75054988; PMID:4215456

A;Accession: A01145
A;Molecule type: protein

A;Residues: 1-259 <TAN>

R;Mallet, B.; Julian, J.M.; Sciaiky, M.; Laurent, G.; Charrel, M.

Biochim. Biophys. Acta 576, 290-304, 1979

C;Keywords: acetylated amino end; carbon-oxygen lyase; hydro-lyase; zinc

F;4-28/Domain: carbonic anhydrase homology <CAH>

P;1/Modified site: acetylated amino end (Ser) #status experimental

P;93,95,118/Binding site: zinc (His) #status predicted

A;Contents: annotation

Qy 6 WGYREHNGPIHWKEFFPIADGDOQSPLEIKTKVYDSSLRPLSIKYDPPSSAKITNSGH 65

Db 4 WGYGERHNGPIHWKDFPPIADGROSPVDPIDTKAVVPPPAKLALLYQQASRRMTNGH 63

Qy 66 SPPVDFDTENKSVLRGSGPLTGSYRLRQHHLWGSADDHGSEHIVDCVSYAAELHYHWN 125

Db 64 SPPKPSFDSQDAVVKQGPGLTGYRLQFHFWGSSHDQGSSTVDRKYYAELHIVHWN 123

Qy 126 SDKYPSEAAHEDPGIAVLGVFLQIGEPNSOLQKITDTSIKEKGKQTRETFNFDLISL 185

Db 124 T-KYGDCTAAQPDGLAVGVFLQVNDLSSIKTKGKSAJDPNFPSSL 182

Qy 186 LPPSWDWTYPPLSLLTVPPLLESVTWILKQPINISSQQLAKERSLCTAEGE 237

Db 183 LKRALNWTYPPLSLTNPLLESVTWVILKEPTSVSSQMLKERSLNFAEGS 234

A;Contents: annotation; X-ray crystallography, 2.0 angstroms, residues 4-259
 R;Liljas, A.; Kannan, K.K.; Bergsten, P.C.; Waara, I.; Fridborg, K.; Strandberg, B.; Car
 Nature New Biol. 235, 131-137, 1972
 A;Title: Crystal structure of human carbonic anhydrase C.
 A;Reference number: A93404; MUID:72111787; PMID:4621826
 A;Contents: annotation; X-ray crystallography, 2.0 angstroms
 A;Note: other residues at the active site are His-64, Asn-67, Tyr-127, Leu-197, Thr-198,
 C;Genetics:
 A;Gene: GDI; CA2
 A;Cross-references: GDB:119739; OMIM:259730
 A;Map position: 8q13-8q22.1
 A;Intron: 12/1; 78/1; 117/3; 148/3; 169/3; 221/3
 C;Keywords: catalyzes the reversible dissociation of carbonic acid to carbon dioxide
 C;Function: carbonic anhydrase homology
 F;5-259/Domain: carbonic anhydrase homology <CAH>
 F;2/Modified site: acetylated amino end (Ser) (in mature form) #status experimental
 F;94,95,119/Binding site: zinc (His) #status experimental
 Query Match 61.2% Score 788.5; DB 1; Length 260;
 Best Local Similarity 61.6%; Pred. No. 2.6e-59; Mismatches 33; Indels 1; Gaps 1;
 Matches 143; Conservative 33; Mismatches 55; Indels 1; Gaps 1;
 Qy 6 WGYREHNGPIHKKEFFPPIADGDDQSPPIEIKTKVYDSSRLPLSIKYDPSAKILISNSGH 65
 Db 5 WGYGKHEKGPEHWEKHDPEPAKGERQSPVDDHTAKYDPSLKLPSVSYQATSLRILNGH 64
 Qy 66 SENVDFDTENKSVLRGGLPTGSYRLRQVHFLWGSADHGSEHIVDGWSYAAELHVWVN 125
 Db 65 AENVEFDQSDKAVALKGGLDGTLLQFHRSQLDQGSEETVDKKYAAELHVWVN 124
 Qy 126 SDKYPSEAAHEDPGIAGLAVLGFLQIGENPSOLQKITTDLDSIKEKGKOTRFNEDLIL 185
 Db 125 T-KYGDFTGAVQQDGLAVLGFLQKGVQDLSITKTKGSSADFTNEDPGL 183
 Qy 186 LPPSWDWTPGSLTVPPPLESIVWIKLQPINISSQQLAKERSLICTAEGE 237
 Db 184 LPESLDWTYPGSLTTTTPPLECIVWIKLKEPIVSSEQVLFKFRXLNFGEGE 235

RESULT 7
 CR022
 carbonate dehydratase (EC 4.2.1.1) II - bovine (tentative sequence)
 N;Alternate names: carbonic anhydrase II
 C;Species: Bos primigenius taurus (cattle)
 C;Date: 31-Mar-1981 #sequence_revision 31-Mar-1981 #text_change 31-Mar-2000
 C;Accession: A01144
 R;Sciaky, M.; Limozin, N.; Filippi-Foreau, D.; Guilian, J.M.; Laurent-Tabusse, G.
 Biochimie 58, 1071-1082, 1976
 A;Title: Structure primaire de l'anhydrase carbonique erythrocytaire bovine CI. II. - Se
 A;Reference number: A90669; MUID:77063798; PMID:826282
 A;Accession: A01144
 A;Molecule type: protein
 A;Residues: 1-259 <SC1>
 R;Guilan, J.M.; Limozin, N.; Mallet, B.; Di Costanzo, J.; Charrel, M.
 Biochimie 59, 293-302, 1977
 A;Title: Independance genetique de deux formes de l'anhydrase carbonique erythrocytaire
 A;Reference number: A90672; MUID:77242599; PMID:19093
 A;Contents: annotation
 A;Note: one minor and two major forms were isolated chromatographically. One of the major
 C;Superfamily: carbonate dehydratase; carbonic anhydrase homology
 C;Keywords: acetylated amino end; carbon-oxygen lyase; hydro-lyase; zinc
 F;4/258/Domain: carbonic anhydrase homology <CAH>
 F;1/Modified site: acetylated amino end (Ser) #status experimental
 F;93,95,118/Binding site: zinc (His) #status predicted

Query Match 60.9% Score 785.5; DB 1; Length 259;
 Best Local Similarity 61.2%; Pred. No. 4.7e-59; Mismatches 41; Indels 1; Gaps 1;

Qy 6 WGYREHNGPIHKKEFFPPIADGDDQSPPIEIKTKVYDSSRLPLSIKYDPSAKILISNSGH 65
 Db 4 WGYGKHEKGPEHWEKHDPEPAKGERQSPVDDHTAKYDPSLKLPSVSYQATSLRILNGH 63
 Qy 66 SENVDFDTENKSVLRGGLPTGSYRLRQVHFLWGSADHGSEHIVDGWSYAAELHVWVN 125
 Db 64 SENVYEDSDQDKAVLKGGLDGTLLQFHRSQLDQGSEETVDKKYAAELHVWVN 123
 Qy 126 SDKYPSEAAHEDPGIAGLAVLGFLQIGENPSOLQKITTDLDSIKEKGKOTRFNEDLIL 185
 Db 124 T-KYGDFTGAVQQDGLAVLGFLQKGVQDLSITKTKGSSADFTNEDPGL 182
 Qy 186 LPPSWDWTPGSLTVPPPLESIVWIKLQPINISSQQLAKERSLICTAEGE 237
 Db 183 LPESLDWTYPGSLTTTTPPLECIVWIKLKEPIVSSEQVLFKFRXLNFGEGE 235
 RESULT 8
 T08463
 carbonate dehydratase (EC 4.2.1.1) - zebra fish
 C;Species: Brachydanio rerio (zebra fish)
 C;Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 13-Aug-1999
 C;Accession: T08463
 R;Peterson, R.E.; Tu, C.; Linser, P.J.
 J. Mol. Evol. 44, 432-439, 1997
 A;Title: Isolation and characterization of a carbonic anhydrase homologue from the zebra fish
 A;Reference number: Z16422; MUID:909083
 A;Accession: T08463
 A;Status: preliminary; translated from GB/EMBL/DBJ
 A;Molecule type: mRNA
 A;Residues: 1-260 <PEP>
 C;Cross references: EMBL:U55177; NID:92576334; PID:92576335
 C;Superfamily: carbonate dehydratase; carbonic anhydrase homology
 C;Keywords: carbon-oxygen lyase; hydro-lyase; zinc
 F;5-259/Domain: carbonic anhydrase homology <CAH>
 Query Match 60.8% Score 783.5; DB 2; Length 260;
 Best Local Similarity 61.8%; Pred. No. 7e-59; Mismatches 54; Indels 1; Gaps 1;
 Matches 144; Conservative 34; Mismatches 54;
 Db 4 AWG3GPADGPESWAESEFPLANGPQSPDIPVQAOHQDPSLKHILKLYDPATPKSILNG 63
 Qy 5 SWGYREHNGPIHKKEFFPPIADGDDQSPPIEIKTKVYDSSRLPLSIKYDPSAKILISNSGH 64
 Db 4 AWG3GPADGPESWAESEFPLANGPQSPDIPVQAOHQDPSLKHILKLYDPATPKSILNG 63
 Qy 65 HSPNVDFFDTENKSVLRGGLPTGSYRLRQVHFLWGSADHGSEHIVDGWSYAAELHVWVN 124
 Db 64 HSPNVDFFDTENKSVLRGGLPTGSYRLRQVHFLWGSADHGSEHIVDGWSYAAELHVWVN 123
 Qy 125 NSDKYPSFVEAAHEDPGIAGLAVLGFLQIGENPSOLQKITTDLDSIKEKGKOTRFNEDLIL 184
 Db 124 NT KYPNFGEEASPDGLAVVGVLKIGANPRIQKVLDLDDKSKGRQTTTANFDPTK 182
 Qy 185 LPPSWDWTPGSLTVPPPLESIVWIKLQPINISSQQLAKERSLICTAEGE 237
 Db 183 LPESLDWTYPGSLTTTTPPLECIVWIKLKEPIVSSEQVLFKFRXLNFGEGE 235
 RESULT 9
 CRM01R
 carbonate dehydratase (EC 4.2.1.1) I - rhesus macaque (tentative sequence)
 N;Alternate name: carbonic anhydrase I
 C;Species: Macaca mulatta (rhesus macaque)
 C;Date: 07-May-1981 #sequence_revision 07-May-1981 #text_change 31-Mar-2000
 C;Accession: A01139
 R;Henriksson, D.; Tanis, R.J.; Tashian, R.E.
 Biochem. Biophys. Res. Commun. 96, 135-142, 1980
 A;Title: The amino acid sequence of carbonic anhydrase I from the Rhesus macaque.
 A;Reference number: A01139; MUID:81062409; PMID:6776950
 A;Accession: A01139
 A;Molecule type: protein
 A;Residues: 1-260 <SHEN>
 C;Superfamily: carbonate dehydratase; carbonic anhydrase homology
 C;Keywords: acetylated amino end; carbon-oxygen lyase; hydro-lyase; zinc

F;5-260/Domain: carbonic anhydrase homology <CAH>
 F;1/Modified site: acetylated amino end (Ala) #status experimental
 F;1/Modified site: zinc (His) #status predicted

Query Match 60.5%; Score 780; DB 1; Length 260;
 Best Local Similarity 58.9%; Pred. No. 1.4e-59;
 Matches 136; Conservative 46; Mismatches 49; Indels 0; Gaps 0;

Qy 6 WGYREHNGPIHKKEFFPPIADGQSPPIEIKRKEVKKDSSLRPLSIKRDPSAKIISNSGH 65
 Db 5 WGYDDKXNGPEQWPSKLKLYTANGNNNOSPVDIKTSSEAKHTSLRPLSIVSYNPAKEINVGH 64
 F;5-259/Domain: carbonic anhydrase homology <CAH>
 F;94, 96, 119/Binding site: zinc (His) #status predicted

Qy 66 SFNVDFFDTENKSVLRGPLTOSYRQVHLWGSADDHGSHFHVQDVSYAELHYHWN 125
 Db 65 SFHVNFDNDNSVLRKGSPFSSPSYRLQFHFWGSNEYGSHTHVQWVYKSELEHIVHN 124
 F;94, 96, 119/Binding site: zinc (His) #status predicted

Qy 126 SDKYPFVEAAHEPDGLAVLGFLQICEPNSLOQKTDIISIKEKGKQTRETNFEDLISL 185
 Db 125 SAKYSSLAEVASKADGIAVIGULMKVSEANPKLQKVDAHLAIKTKGRAPFTNFEDPL 184
 F;94, 96, 119/Binding site: zinc (His) #status predicted

Qy 186 LPPSWDYWTYPPGSLTYPPPLESTVWVTKOPINISSQLAKEFSLCTAG 236
 Db 185 LBSSLDFWTYPSLTHPPLYESTVWTCKESSVSQQLAQFRSLLSNVEG 235
 F;94, 96, 119/Binding site: zinc (His) #status predicted

RESULT 10
 JC2580
 A;Reference number: JC2580; MUID:95011620; PMID:7926806
 A;Accession: JC2580
 A;Molecule type: mRNA
 A;Residues: 1-260 <PER>
 A;Cross-references: EMBL:Z14957; NID:965331; PIDN:CAA78681.1; PMID:965332
 C;Species: Gallus gallus (chicken)
 C;Date: 13-Jun-1995 #sequence_revision 14-Jul-1995 #text_change 20-Jun-2000
 R;Goodbou, R.; Andison, R.; Upton, C.; Day, R.
 Nucleic Acids Res. 18, 1049, 1990
 A;Title: Utilization of the second polyadenylation signal at the 3' end of the chicken oocyte mRNA
 A;Reference number: S10229; MUID:90192090; PMID:1569140
 A;Accession: S10229
 A;Molecule type: mRNA
 A;Residues: 222-260 <GOD>
 A;Cross-references: EMBL:X17378; NID:961127; PIDN:CAA35250.1; PMID:963128
 R;Yoshihara, C.M.; Lee, J.D.; Dodgson, J.B.
 Nucleic Acids Res. 15, 753-770, 1987
 A;Title: The chicken carbonic anhydrase II gene: evidence for a recent shift in intron F
 A;Reference number: S01078; MUID:87146391; PMID:3029691
 A;Accession: S01078
 A;Molecule type: DNA
 A;Residues: 1-219, V-251-260 <YOS>
 A;Cross-references: EMBL:X06000; NID:963115; PIDN:CAA22417.1; PID:91289219
 A;Note: the authors translated the codon GTA for residue 250 as Leu
 R;Rogers, J.H.
 Eur. J. Biochem. 162, 119-122, 1987
 A;Residues: 1, 6-260 <ROG>
 A;Reference number: A2641; MUID:87333522; PMID:3102231
 A;Accession: A26415
 A;Molecule type: mRNA
 A;Residues: 1, 6-260 <ROG>
 A;Cross-references: EMBL:X04810; NID:963129; PIDN:CAA28501.1; PMID:9833606
 R;Yoshihara, C.M.; Federici, M.; Dodgson, J.B.
 Ann. N. Y. Acad. Sci. 429, 332-334, 1984
 A;Title: Isolation of the chicken carbonic anhydrase II gene.
 A;Reference number: I50181; MUID:84255154; PMID:6331256
 A;Accession: I50181

RESULT 11
 CRRB;
 A;Reference number: A01142; MUID:78144871; PMID:416851
 A;Accession: A01142
 A;Molecule type: protein
 A;Residues: 1-259 <PER>
 A;Note: 20-Glu was also found
 A;Superfamily: carbonic anhydrase homology
 C;Species: Oryctolagus cuniculus (domestic rabbit)
 C;Date: 31-Mar-1981 #sequence_revision 31-Mar-1981 #text_change 31-Mar-2000
 C;Accession: A01142
 R;Perrell, R.E.; Stroup, S.K.; Tanis, R.J.; Tashian, R.E.
 Biochem. Biophys. Acta 533, 1-11, 1978
 A;Title: Amino acid sequence of rabbit carbonic anhydrase II.
 A;Reference number: A01142; MUID:78144871; PMID:416851
 A;Accession: A01142
 A;Molecule type: protein
 A;Residues: 1-259 <PER>
 A;Superfamily: carbonic anhydrase homology
 C;Keywords: acetylated amino end; carbon-oxygen lyase; hydro-lyase; zinc
 F;4-58/Domain: carbonic anhydrase homology <CAH>
 F;93-94/Binding site: acetylated amino end (Ser) #status experimental
 F;93, 95, 118/Binding site: zinc (His) #status predicted

Query Match 59.1%; Score 761.5; DB 1; Length 259;
 Best Local Similarity 60.8%; Pred. No. 5.1e-57;
 Matches 141; Conservative 31; Mismatches 59; Indels 1; Gaps 1;

Qy 6 WGYREHNGPIHKKEFFPPIADGQSPPIEIKRKEVKKDSSLRPLSIKRDPSAKIISNSGH 65
 Db 4 WGYKINGPHEWPKDPIADGERQSIIDDTAAKHDPSLKPLRVSEHPSRSRRTNNGH 63
 F;94, 96, 119/Binding site: zinc (His) #status predicted

Qy 66 SFNVDFFDTENKSVLRGPLTOSYRQVHLWGSADDHGSHFHVQDVSYAELHYHWN 125
 Db 64 SFNVEEDDSHDKSVLKPLGVPLETRIQLQFHFWGSSDGESEEHVVKKKYAAELHVHN 123
 F;94, 96, 119/Binding site: zinc (His) #status predicted

Qy 126 SDKYPFVEAAHEPDGLAVLGFLQICEPNSLOQKTDIISIKEKGKQTRETNFEDLISL 185
 Db 124 T-KYGFKAIVHPDGLAVGIFLKGQSATPGLQKVDTLSSIKTKGKSYDFTNFDPRGL 182
 F;94, 96, 119/Binding site: zinc (His) #status predicted

Qy 186 LPPSWDYWTYPPGSLTYPPPLESTVWVTKOPINISSQLAKEFSLCTAG 236

| | | | | | |
|--|-------|---|-------|--|--|
| Db | 183 | LPESDLQWYTPGSLITTPPLIQCVCYTIVLKEPITVSSSEQMLKFRNLNFNKEAE | 234 | | Best Local Similarity 59.2%; Pred. No. 2.e-56; Matches 142; Conservative 36; Mismatches 60; Indels 2; Gaps 2; |
| RESULT 12 | | | | | |
| A22612 | | carbonate dehydratase (EC 4.2.1.1) III - horse | | Qy 1 MSRLSMCGYREHFINGPHWKEFFPPIADQDQSPLEIKTKVEVKYDSSRPLSKYDPSSAKII 60 | |
| N;Alternative names: carbonate dehydratase III | | C;Species: Equus caballus (domestic horse) | | Db 1 MASADMGYGSENGPDOWSKLYPIANGMNQSPIDKTEAHNDSSLKPLSISYNPAKEI 60 | |
| C;Date: 22-Jul-1987 #sequence_revision 22-Jul-1987 #text_change 25-Apr-1997 | | R;Wendorff, K.M.; Nishita, T.; Jabusch, J.R.; Deutsch, H.F. | | Qy 61 SNSGHSHFNVDDTENKSVLVLRGGLPTGSYRLRQVHLHWG-SADDGSEHITYDGVSAYAAEL 119 | |
| R;Wendorff, K.M.; Nishita, T.; Jabusch, J.R.; Deutsch, H.F. | | J. Biol. Chem. 260, 6129-6132, 1985 | | Db 61 VNGHSHFVIPEDSSNSVVLK-GLPLADSYRQJOFHHWGNNSNDSESEHTYDGTRYSGEL 119 | |
| A;Title: The sequence of equine muscle carbonic anhydrase. | | A;Reference number: A22612; MUID:85207593; PMID:3922970 | | Qy 120 HVVHWNSDKPSFEVAHEPDGLAVLGVELQIGEPNSQLOKTTDLDLSIKEKGKQTREFTN 179 | |
| A;Molecule type: protein | | A;Accession: A22612 | | Db 120 HLVHWNSAKYSSASEAISKAISGCLALGVLMKVGPANSLQKVLDALNSVKTGKRAFTN 179 | |
| A;Residues: 1-259 <WEN> | | | | Qy 180 FDLLSLPPSWDYTWYTPGSILTVPPPLESVTVLKKOPINISSQQQLAKFRSLCTACEAA 239 | |
| C;Keywords: acetylated amino end; carbonate dehydratase; carbonic anhydrase homology <CAH> | | | | Db 180 FDPSLIPSLPSSLDYTWYTFEGSLATHPPLHESVTWVICKDOSISLSPFQLAQRLGLISSAEESA 239 | |
| F;4-258/Domain: carbonic anhydrase homology <CAH> | | | | | |
| F;1/Modified site: acetylated amino end (Ala) #status predicted | | | | | |
| Query Match 58.5%; Score 754.5%; DB 2; Length 259; | | | | | RESULT 14 |
| Best Local Similarity 59.1%; Pred. No. 2.e-56; | | | | | CRMS2 |
| Matches 137; Conservative 30; Mismatches 64; Indels 1; Gaps 1; | | | | | carbonic dehydratase (EC 4.2.1.1) II - mouse |
| Qy 6 WGCREHNSPHWKEFFPPIADQDQSPLEIKTKVEVKYDSSRPLSKYDPSSAKII 65 | | | | | N;Alternative names: carbonic anhydrase II |
| Db 4 WGYADHNGPDPDWHEFYPPAKGDNQSPLEHTKDNHDSLKAWTASDPGSAKTLNNGR 63 | | | | | C;Species: Mus musculus (house mouse) |
| Qy 66 SPTVDFDGTENKSVLVLRGGLPTGSYRLRQVHLHWGSDADHGSEHITYDGVSAYAAELHVWN 125 | | | | | C;Date: 25-Feb-1985 #sequence_revision 06-Feb-1995 #text_change 18-Jun-1999 |
| Db 64 TCKWVFDITYDRSMLRGSPLTAYRLRQPHLAWGSSDHGSEITVDGKYAELHVWN 123 | | | | | R;Venta, P.J.; Montgomery, C.; Hewett-Bennett, D.; Tashian, R.E. |
| Qy 126 SDKXPSFVAAHEPDGLAVLGVLQIGEPNSQLOKTTDLSIKEKGKQTREFTNFDLSSL 185 | | | | | A;Accession: A23900; B23202; A01143; PMID:10339; I51949 |
| Db 124 -PKNTYCGALKQPDGIVKIGERKQFLKTKTGEAPFTNFEDPSCL 182 | | | | | R;Venta, P.J.; Montgomery, J.C.; Hewett-Bennett, D.; Wiebauer, K.; Tashian, R.E. |
| Qy 186 LPTSDWYTPGSILTVPPPLESVTVLKKOPINISSQQQLAKFRSLCTACEE 237 | | | | | J. Biol. Chem. 260, 12130-12135, 1985 |
| Db 183 FPTCRDWTYRGSEFTTPCECIVWLLKEPITVSSDQYAKLRSLSFSSAENE 234 | | | | | A;Title: Comparison of the 5' regions of human and mouse carbonic anhydrase II genes and exon to protein domain relationships of the mouse carbonic anhydrase II genes |
| | | | | | A;Reference number: A23900; PMID:2995362 |
| | | | | | A;Accession: A23900; PMID:86008276; PMID:2995362 |
| | | | | | A;Molecule type: DNA |
| | | | | | A;Residues: 1-260 <VEN> |
| | | | | | A;Experimental source: strain YBR |
| | | | | | R;Venta, P.J.; Montgomery, C.; Hewett-Bennett, D.; Tashian, R.E. |
| | | | | | Biochim. Biophys. Acta 826, 195-201, 1985 |
| | | | | | A;Title: Structure and exon to protein domain relationships of the mouse carbonic anhydrase II genes |
| | | | | | A;Reference number: A90655; PMID:86077780; PMID:3000449 |
| | | | | | A;Accession: B23202 |
| | | | | | A;Molecule type: DNA |
| | | | | | A;Residues: 1-77 <VE2> |
| | | | | | A;Note: the authors translated the codon CAG for residue 39 as His |
| | | | | | R;Curtis, P.J.; Withers, E.; Demuth, D.; Watt, R.; Venta, P.J.; Tashian, R.E. |
| | | | | | Gene 25, 325-332, 1983 |
| | | | | | A;Title: The nucleotide sequence and derived amino acid sequence of cDNA coding for mouse carbonic anhydrase II |
| | | | | | A;Reference number: A01143; PMID:6420240 |
| | | | | | A;Accession: A01143 |
| | | | | | A;Molecule type: mRNA |
| | | | | | A;Cross references: GB:K00911; GB:K00812; GB:ML1830; PID:9192333; PID:AAA37356.1; PID:9192333 |
| | | | | | A;Note: initiator Met not shown |
| | | | | | R;Curtis, P.J. |
| | | | | | A;Title: Cloning of mouse carbonic anhydrase mRNA and its induction in mouse erythroleukemic cells |
| | | | | | A;Reference number: A20539; PMID:83161023; PMID:6187736 |
| | | | | | A;Accession: A20539 |
| | | | | | A;Molecule type: mRNA |
| | | | | | A;Residues: 1-55-178; 214-240 <CU2> |
| | | | | | R;Venta, P.J.; Montgomery, J.C.; Wiebauer, K.; Hewett-Bennett, D.; Tashian, R.E. |
| | | | | | Ann. N. Y. Acad. Sci. 429, 309-323, 1984 |
| | | | | | A;Title: Organization of the mouse and human carbonic anhydrase II genes |
| | | | | | A;Reference number: I51949; PMID:8425152; PMID:6331255 |
| | | | | | A;Accession: I51949 |
| | | | | | A;Status: preliminary; translated from GB/EMBL/DDBJ |
| | | | | | A;Cross references: GB:M25944; PMID:8425152 |
| | | | | | A;Molecule type: mRNA |
| | | | | | A;Residues: 241-260 <RES> |
| | | | | | C;Genetics: |
| | | | | | A;Gene: Car-2 |
| | | | | | A;Introns: 12/1; 78/1; 117/3; 144/1; 169/3; 221/3 |
| | | | | | C;Superfamily: carbonate dehydratase; carbonic anhydrase homology |
| | | | | | F;6-261/Domain: carbonic anhydrase homology <CAH> |
| Query Match | 58.5% | Score 754; | DB 2; | Length 261; | |

P;5-259/Domain: carbonic anhydrase homology <CAH>
F;94,96,119/Binding site: zinc (His) #status predicted

Query Match 58.1%; Score 749.5; DB 1; Length 260;
Best Local Similarity 58.4%; Pred. No. 5.3e-56;
Matches 136; Conservative 42; Mismatches 54; Indels 1; Gaps 1;
Qy 6 WGYREHNGPIHWKEFFPPIADGDDQSPPIEIKTKVEVKYDSSURPLSITKDPSSAKIISNSGH
Db 5 WGYSSRNGPENWHEKFPIANGRQSPVDDIAQAQIDPAQELLSYDKAASKSTVNNGH 64
Qy 66 SENVDDDTENKSVLRGGLTGSYRLRQVHLWWSADDHGSEHTIVDGVSAAELHVHN 125
Db 65 SENVFDDSDQNAVLRKGELPSDYSYRLQFHFHWGSSDGOQSEHTVNKKVAAELHVHN 124
Qy 126 SDKYPSEVERAHEPPGLAVLGFLQIGEPNSLQLKTTDLDTSIKEKGKQTRFTNFDLISL 185
Db 125 T-KYGFGEKAQVQDPGLAVLGFLKPGASQGLQKVLEALHSIKTKGKRAAFANFDPCSL 183
Qy 186 LPPSWDYWTVPGLSLTVPPLESVTWVLKOPINISSQQLAKFRSLICCTAGEA 238
Db 184 LGPNLDWYTVPGLSLTVPPLESVTWVLREPTVSVLRLPECTVTLRPPLECVTIVLREFITVSSEQMHSERTLNNEEGDA 236

RESULT 15

carboxane dehydratase (EC 4.2.1.1) II - rat
N;Alternate names: carbonic anhydrase II

C;Species: Rattus norvegicus (Norway rat)

C;Date: 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change 22-Jun-1999

C;Accession: JH0527; S2014

R;Stoile, C.A.; McGowan, M.H.; Heim, R.A.; Varia, M.; Neubauer, J.A.

Gene 109, 265-267, 1991

A;Title: Nucleotide sequence of a cDNA encoding rat brain carbonic anhydrase II and its

A;Reference number: JH0527; MUID: 92112053; PMID: 1765271

A;Accession: JH0527

A;Molecule type: mRNA

A;Residues: 1-260 <STO>

A;Cross-References: EMBL:X58294; NID:955837; PIDN:CAA41227.1; PID:955838

A;Experimental source: brain

C;Comment: Carbonate dehydratase is a monomeric zinc metalloenzyme that catalyzes the re

C;Superfamily: carbonate dehydratase; carbonic anhydrase homology

C;Keywords: carbon-oxygen lyase; hydro-lyase; zinc

F;5-259/Domain: carbonic anhydrase homology <CAH>

Query Match 58.1%; Score 748.5; DB 2; Length 260;
Best Local Similarity 60.1%; Pred. No. 6.4e-56;
Matches 140; Conservative 37; Mismatches 55; Indels 1; Gaps 1;
Qy 6 WGYREHNGPIHWKEFFPPIADGDDQSPPIEIKTKVEVKYDSSURPLSITKDPSSAKIISNSGH 65
Db 5 WGYSSRNGPENWHEKFPIANGRQSPVDDIAQAQIDPAQELLSYDKAASKSTVNNGH 64
Qy 66 SENVDDDTENKSVLRGGLTGSYRLRQVHLWWSADDHGSEHTIVDGVSAAELHVHN 125
Db 65 SENVFDDSDQNAVLRKGELPSDYSYRLQFHFHWGSSDGOQSEHTVNKKVAAELHVHN 124
Qy 126 SDKYPSEVERAHEPPGLAVLGFLQIGEPNSLQLKTTDLDTSIKEKGKQTRFTNFDLISL 185
Db 125 T-KYGFGEKAQVQDPGLAVLGFLKPGASQGLQKVLEALHSIKTKGKRAAFANFDPCSL 183
Qy 186 LPPSWDYWTVPGLSLTVPPLESVTWVLKOPINISSQQLAKFRSLICCTAGEA 238
Db 184 LGPNLDWYTVPGLSLTVPPLESVTWVLREFITVSSEQMHSERTLNNEEGDA 236

Search completed: September 9, 2004, 15:14:40
Job time : 41 secs

GenCore version 5.1.6
 Copyright (c) 1993 - 2004 Compugen Ltd.
 OM protein - protein search, using sw model
 Run on: September 9, 2004, 14:58:23 ; Search time 124 Seconds
 (without alignments)
 551.423 Million cell updates/sec
 Title: US-10-069-434-1
 Perfect score: 1289
 Sequence: 1 MSRLSWGYREHNGPIHWKEF.....QLAKFERSLLCTAEGEAAFL 242
 Scoring table: BLOSUM62
 Gapop 10.0 , Gapext 0.5
 Searched: 158617 seqs, 282547505 residues
 Total number of hits satisfying chosen parameters: 1586107

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_29Jan04:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Match | Length | ID | DB | Description |
|------------|--------|-------|--------|----|----------|-------------|
| 1 | 1289 | 100.0 | 242 | 5 | AB08900 | Human lya |
| 2 | 1289 | 100.0 | 262 | 6 | ABJ37886 | NOVX prot |
| 3 | 1289 | 100.0 | 274 | 4 | AU19418 | Human dia |
| 4 | 1288 | 99.9 | 262 | 6 | ABJ37885 | NOVX prot |
| 5 | 1193.5 | 92.6 | 247 | 5 | AAB25377 | Human NZM |
| 6 | 864 | 67.0 | 184 | 3 | AAB63110 | Human sec |
| 7 | 804 | 62.4 | 261 | 5 | AAO15236 | Human car |
| 8 | 804 | 62.4 | 261 | 6 | ADA10962 | Human CDN |
| 9 | 804 | 62.4 | 263 | 4 | AAG73863 | Human col |
| 10 | 802 | 62.0 | 260 | 4 | AAB59588 | Human car |
| 11 | 791.5 | 61.4 | 260 | 2 | AAW75702 | Carbonic |
| 12 | 788.5 | 61.2 | 259 | 4 | AAB59589 | Human car |
| 13 | 788.5 | 61.2 | 259 | 7 | ADE62800 | Human Pro |
| 14 | 788.5 | 61.2 | 259 | 7 | ADE63735 | Human Pro |
| 15 | 788.5 | 61.2 | 260 | 6 | ADA10999 | Human cDN |
| 16 | 788.5 | 61.2 | 288 | 2 | AAW23378 | Fusion co |
| 17 | 788.5 | 61.2 | 294 | 3 | AAB53405 | Human col |
| 18 | 761.5 | 59.1 | 259 | 7 | ADE54976 | Rat Prote |
| 19 | 750 | 58.2 | 143 | 4 | AAO12529 | Human pol |
| 20 | 748.5 | 58.1 | 259 | 4 | AAB59590 | Human pol |
| 21 | 748.5 | 58.1 | 259 | 7 | ADE54978 | Human Pro |
| 22 | 748.5 | 58.1 | 259 | 7 | ADE62798 | Rat Prote |
| 23 | 748.5 | 58.1 | 259 | 7 | Ade63133 | Rat Prote |
| 24 | 748.5 | 58.1 | 260 | 5 | ABB57256 | Mouse inc |
| 25 | 747.5 | 58.0 | 260 | 5 | ABP62813 | Human inc |

ALIGNMENTS

RESULT 1
ABB08900 ID ABB08900 standard; protein: 242 AA.

| | | |
|-----------|-------------|---|
| ABB08900; | 18-JUN-2002 | (first entry) |
| | | Human lyase HLYA-1 protein. |
| | | Cystostatic; anticonvulsant; cerebroprotective; nootropic; virucide; neuroprotective; antibacterial; antiidiabetic; antiinflammatory; antigout; ophthalmological; hypotensive; immunosuppressive; dermatological; nephrotrophic; antithyroid; thyromimetic; osteopathic; antipsoriatic; antitulerc; fungicide; antiparasitic; protozocide; tranquiliser; cancer; neuroleptic; diagnosis; treatment; immunological disorder; AIDS; allergy; acquired immunodeficiency syndrome; asthma; HLYA; infection; anaemia; Crohn's disease; multiple sclerosis; atherosclerosis; osteoporosis; stroke; rheumatoid arthritis; Alzheimer's disease; epilepsy; stroke; muscular dystrophy; Down's syndrome; myasthenia gravis; glaucoma; transgenic; gene therapy; drug screening; human lysis; enzyme |

Homo sapiens.

WO200200840-A2.

03-JAN-2002.

13-JUN-2001; 2001WO-US019166.

23-JUN-2000; 2000US-0213383P.

30-JUN-2000; 2000US-0215544P.

04-AUG-2000; 2000US-0222818P.

(INCYT-) INCYTE GENOMICS INC.

Thornton M, Ramkumar J, Tribouley CM, Yue H, Nguyen DB, Yao MG;
Patterson C, Gandhi AR, Burford N, Thangavelu K, Baughn MR;

WPI, 2002-139910/18.

New isolated human lyase polypeptide for diagnosing, treating and preventing e.g. glaucoma, ocular hypertension, stroke, asthma, or gout.

Claim 1; Page 96-97; 101pp; English.

The present invention concerns human lyase polypeptides, human lyase polypeptides having the sequence shown in claim 1, and compositions

by the polynucleotide given in ABA97691. The specification describes an isolated HLYA polypeptide or a nucleic acid that encodes it. The invention has cytoprotective, roottropic, anticonvulsant, cerebroprotective, virucide, neuroprotective, anti-HIV, antiparkinsonian, antibacterial, antidiabetic, antiinflammatory, ophthalmological, hypotensive, antiallergic, antiallergic, antiaemic, antiasthmatic, antihypertrophic, immunosuppressive, dermatological, nephrotoxic, thyromimetic, osteopathic, antipsoriatic, antiarthritic, dermatological, fungicide, antiparasitic, protozoacide, tranquiliser and neuroleptic applications. The protein of the invention may be used to screen for potential HLYA agonists or antagonists; detect the presence of HLYA; associates disorders; assess the toxicity of a test compound. The HLYA proteins and polynucleotides are useful in diagnosis, treatment and prevention of immunological disorders e.g. AIDS, allergy, anaemia, asthma, infection, Crohn's disease, multiple sclerosis, atherosclerosis, rheumatoid arthritis, osteoporosis; cancer, neurological disorder e.g. Alzheimer's and Parkinson's disease, epilepsy, stroke, muscular dystrophy, Down's syndrome, myasthenia gravis; Glaucoma. HLYA polynucleotides are used for creating humanised/transgenic animals to model human diseases; somatic or germline gene therapy; for generating mapping probes for e.g. gene mapping; detecting differences in chromosomal location due to e.g. translocation; Generating a transcript image of a tissue/cell type. Antibodies which bind to the HLYA proteins are used for diagnosis of HLYA-associated disorders or monitoring patients being treated with HLYA or agonists, antagonists or inhibitors of HLYA and for assessing toxicity of a test compound

Sequence 242 AA:
Query Match 100.0%; Score 1289; DB 5; Length 242;
Best Local Similarity 100.0%; Prod. No. 6e-126; Indels 0; Gaps 0;
Matches 242; Conservative 0; Mismatches 0;

QY 1 MSRLSWGKREHNGPTHWKEFFPIADGQOSPIEIKTKKEYKDSSLRPLSIKYDSSAKII 60
1 MSRLSWGKREHNGPTHWKEFFPIADGQOSPIEIKTKKEYKDSSLRPLSIKYDSSAKII 60
DB 61 SNSGHFSFVYDFDDTENKSVLRLGGPITGSYRLROYHLNGSADDHGSEHIVDGYSYAAHL 120
61 SNSGHFSFVYDFDDTENKSVLRLGGPITGSYRLROYHLNGSADDHGSEHIVDGYSYAAHL 120
QY 121 VVHNNSDKYPSFVRAAHEPDGLAVLGVLQIGEPNSQLOQKITDTLSIKEKGKQTREFNF 180
121 VVHNNSDKYPSFVRAAHEPDGLAVLGVLQIGEPNSQLOQKITDTLSIKEKGKQTREFNF 180
DB 181 DLLSLLPSPMDWYWPGSTTVPLLESTWIVLKOPINISSQQLAKFSLCTAEGAAA 240
181 DLLSLLPSPMDWYWPGSTTVPLLESTWIVLKOPINISSQQLAKFSLCTAEGAAA 240
DB 241 FL 242
QY 241 FL 242
DB 241 FL 242

RESULT 2
ID ABJ37886 standard; protein: 262 AA.

XX ABJ37886;
XX DT 22-MAY-2003 (first entry)

KW Parasitic infection; Alzheimer's disease; stroke; forensic biology.
CC KW immunogen; non-human transgenic animal; gene therapy.
CC KW Unidentified.
XX OS WO200291517-A2.
XX PN 17-OCT-2002.
XX PD 22-JAN-2002; 2002WO-US002064.
XX PR 19-JAN-2001; 2001US-0262892P.
CC PR 23-JAN-2001; 2001US-0263598P.
CC PR 24-JAN-2001; 2001US-0263798P.
CC PR 25-JAN-2001; 2001US-0264112P.
CC PR 25-JAN-2001; 2001US-0264139P.
CC PR 26-JAN-2001; 2001US-0264478P.
CC PR 30-JAN-2001; 2001US-0263351P.
CC PR 02-MAR-2001; 2001US-0272870P.
CC PR 14-MAR-2001; 2001US-0275927P.
CC PR 14-MAR-2001; 2001US-0275990P.
CC PR 15-MAR-2001; 2001US-0276449P.
CC PR 20-MAR-2001; 2001US-0277358P.
CC PR 23-MAR-2001; 2001US-0278151P.
CC PR 29-MAR-2001; 2001US-0279857P.
CC PR 20-APR-2001; 2001US-0285140P.
CC PR 20-APR-2001; 2001US-0285141P.
CC PR 30-APR-2001; 2001US-0287484P.
CC PR 17-MAY-2001; 2001US-0291701P.
CC PR 08-JUN-2001; 2001US-0296960P.
CC PR 10-JUL-2001; 2001US-0304353P.
CC PR 12-JUL-2001; 2001US-0304355P.
CC PR 09-AUG-2001; 2001US-0311289P.
CC PR 16-AUG-2001; 2001US-0311975P.
CC PR 16-AUG-2001; 2001US-0312937P.
CC PR 18-OCT-2001; 2001US-0330227P.
CC PR 29-NOV-2001; 2001US-0334198P.
XX PA (CTRA-) CURAGEN CORP.
XX DR N-PSDB; ABT3351.
XX PT New polypeptides, designated as NOVX, useful for diagnosing and treating
PT immunological, skin, renal, brain, muscle and autoimmune disorders.
XX

Claim 1; Page 67; 672pp; English.

CC The invention relates to a novel isolated polypeptide, designated NOVX (NOV1 - 33), consisting of a mature form of one of 61 sequences, given in the specification, or its variant, where amino acid residue(s) in the variant differ from the mature form, provided that the variant differs in not more than 15 % of the amino acids from the sequence of the mature form. The NOVX polypeptides, nucleic acids encoding the polypeptides, and an antibody to the polypeptides, are useful for treating or preventing a NOVX-associated disorder in humans and for treating a syndrome associated with a human disease (NOVX-associated disorder). NOVX polypeptides and the encoding nucleic acids are useful for determining the presence of or predisposition to a disease associated with altered levels of NOVX polypeptide and polynucleotide, by measuring the level of polypeptide expression or the amount of nucleic acid from a mammal and comparing it with another mammal not having or not predisposed to the disease. NOVX polypeptide is also useful for identifying an agent that binds to NOVX and a cell expressing NOVX is useful for identifying an agent that

modulates the expression or activity of NOVX. The antibodies and a polypeptide having 95 % sequence identity to NOVX polypeptide are useful for treating a pathological state in a mammal. The antibodies are also useful for determining the presence or amount of NOVX in a sample. NOVX polypeptides, polynucleotides and antibodies specific for the polypeptides are useful for treating or preventing disorders or syndromes including trauma, viral, bacterial, fungal, protozoal, and parasitic infections. They can also treat disorders such as e.g., Alzheimer's disease or a stroke. The NOVX encoding nucleic acids are useful for expressing the NOVX protein, to detect NOVX mRNA, or a genetic lesion in a NOVX gene and to modulate NOVX activity. NOVX sequences are also useful for identifying a cell or tissue type in a biological sample, to amplify DNA sequences from very small biological samples such as tissues e.g., hair or skin or body fluids in forensic biology and as primers and probes for use in identifying and/or cloning NOVX homologues in other cell types. The NOVX proteins are useful as an immunogen to generate antibodies which are useful for diagnostically monitoring protein levels and modulating NOVX activity. Cells comprising NOVX nucleic acids are useful for producing non-human transgenic animals which are useful for studying the function and/or activity of NOVX protein and for identifying and/or evaluating modulators of NOVX protein activity. The NOVX nucleic acids can be used in gene therapy. This sequence represents a NOVX protein of the invention.

Sequence 262 AA;

Query Match 100.0%; Score 1289; DB 6; Length 262;
Best Local Similarity 100.0%; Pred. No. 6.8e-126;
Matches 242; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

| | | |
|----|-----|--|
| Qy | 1 | MSPLWSGYREHNPIMKEFFPITADGDDQSPIBIKTKKVYKDSSLRPLSIKYDPPSAKII 60 |
| Db | 1 | MSPLWSGYREHNPIMKEFFPITADGDDQSPIBIKTKKVYKDSSLRPLSIKYDPPSAKII 60 |
| Qy | 61 | SNSGHSPNVDFFDTTENKSVLRLGGPLTSYRQLRQVHLWGSADDHGSEHIVDGVSYAELH 120 |
| Db | 61 | SNSGHSPNVDFFDTTENKSVLRLGGPLTSYRQLRQVHLWGSADDHGSEHIVDGVSYAELH 120 |
| Qy | 121 | VVHNNSDKYPSVEAAHEPDGLAVLGFLQCEPNSLQKIDTDLNSIKEKGKQTFTNF 180 |
| Db | 121 | VVHNNSDKYPSVEAAHEPDGLAVLGFLQCEPNSLQKIDTDLNSIKEKGKQTFTNF 180 |
| Qy | 181 | DLSLIPPSWDWVTPSSLTVPPLLESVTWVLYKOPENISQSLAKERSLLCTAEGAAA 240 |
| Db | 181 | DLSLIPPSWDWVTPSSLTVPPLLESVTWVLYKOPENISQSLAKERSLLCTAEGAAA 240 |
| Qy | 241 | FL 242 |
| Db | 241 | FL 242 |

SQ

XX PR 24-FEB-2000; 20000US-0184693P.
CC PR 24-FEB-2000; 20000US-0184693P.
CC PR 24-FEB-2000; 20000US-0184693P.
CC PR 24-FEB-2000; 20000US-0184768P.
CC PR 24-FEB-2000; 20000US-0184769P.
CC PR 24-FEB-2000; 20000US-0184770P.
CC PR 24-FEB-2000; 20000US-0184771P.
CC PR 24-FEB-2000; 20000US-0184772P.
CC PR 24-FEB-2000; 20000US-0184773P.
CC PR 24-FEB-2000; 20000US-0184774P.
CC PR 24-FEB-2000; 20000US-0184776P.
CC PR 24-FEB-2000; 20000US-0184777P.
CC PR 24-FEB-2000; 20000US-0184778P.
CC PR 12-MAY-2000; 20000US-020385P.
CC PR 15-MAY-2000; 20000US-020426P.
CC PR 16-MAY-2000; 20000US-020425P.
CC PR 16-MAY-2000; 20000US-0204321P.
CC PR 16-MAY-2000; 20000US-020432P.
CC PR 17-MAY-2000; 20000US-0204815P.
CC PR 17-MAY-2000; 20000US-020483P.
CC PR 17-MAY-2000; 20000US-0205221P.
CC PR 17-MAY-2000; 20000US-0205285P.
CC PR 17-MAY-2000; 20000US-0205286P.
CC PR 17-MAY-2000; 20000US-0205287P.
XX PA (INCYTE GENOMICS INC.

XX PI Panzer SR, Spiro PA, Banville SC, Shah P, Chalup MS, Chang SC; Chen A, D' sa SA, Amshey S, Dahl CR, Dam TC, Daniels SE, Dufour GE; Flores V, Fong WT, Greenawalt LB, Hillman JL, Jones AL, Liu TF; Roseberry AM, Rosen BH, Russo FD, Stockdher TK, Daffo A; Wright RJ, Yap PE, Yu JY, Bradley DL, Bratcher SR, Chen W; Cohen HJ, Hodson DM, Lincoln SE, Jackson S; DR WPI; 2001-502867/55. DR N-PSDB; ARS3099.

XX PT Polynucleotides encoding diagnostic and therapeutic proteins, e.g., enzymes, hormones and receptors, useful in diagnostics and therapeutics.

XX PS Claim 27; Page 399-400; 522pp; English.

XX CC The invention relates to polynucleotides (I) encoding diagnostic and therapeutic (DTHP) polypeptides (II), which include e.g. enzymes, and proteins involved in growth and development and receptors. (I) and (II) may be used in the prevention, diagnosis and treatment of diseases associated with inappropriate DTHP expression. For example, (I) and (II) may be used to treat disorders associated with decreased polypeptide expression by rectifying mutations or deletions in a patient's genome, that affect the activity of the DTHPs, by expressing inactive proteins or supplementing the patient's own production of them. (I) and (II) may be used to treat diseases, for example, cell proliferative disorder, Crohn's disease, acquired immune deficiency syndrome (AIDS), lymphoma, leukaemia, autoimmune disorders, and respiratory disorders. Additionally, (I) may be used to produce the DTHPs, by inserting the nucleic acids into a host cell and culturing the cell to express the protein. (I) and its complementary sequences may also be used as DNA probes in diagnostic assays to detect and quantitate the presence of similar nucleic acids in samples, and therefore which patients may be in need of restorative therapy. (II) may also be used as antigens in the production of antibodies against DTHPs and in assays to identify modulators of DTHP expression and activity. The anti-DTHP antibodies and antagonists may also be used to down regulate expression and activity. The anti-DTHP antibodies may also be used as diagnostic agents for detecting the

XX DT 04-DEC-2001 (first entry)

XX DE Human diagnostic and therapeutic polypeptide (DTHP) #4.

XX ID AAU19418 standard; protein; 274 AA.

XX AC AAU19418;

XX DT 04-DEC-2001 (first entry)

XX DE Human receptor; diagnostic; therapeutic; gene therapy; vaccine; cell proliferative disorder; Crohn's disease; lymphoma; leukaemia; acquired immune deficiency syndrome; AIDS; autoimmune disorder; respiratory disorder.

XX OS Homo sapiens.

XX PN WO200162927-A2.

XX PD 30-AUG-2001.

XX PF 21-FEB-2001; 2001WO-US006059.

CC presence of DITHP in samples (e.g., by enzyme linked immunosorbant assay
 CC (ELISA)). AAU1915-AAU1965 represent human diagnostic and therapeutic
 CC (DITHP) polypeptides of the invention.

SQ Sequence 274 AA:

Query Match 100.0%; Score 1289; DB 4; Length 274;
 Best Local Similarity 100.0%; Pred. No. 7.3e-12.6; Indels 0; Gaps 0;
 Matches 242; Conservative 0; Mismatches 0;

Qy 1 MSRLSWGYTREHNGPIHWKEFFPPIADGQQSPLIEBKTKVYKDSSIRLPSIKYDPSSAKII 60

Db 13 MSRLSWGYTREHNGPIHWKEFFPPIADGQQSPLIEBKTKVYKDSSIRLPSIKYDPSSAKII 72

Qy 61 SNSGHSSFWVDFDDTENKSVLRGGPLTSYRQLQVHLSWGSADDHGSHIIVDGVSYAAELH 120

Db 73 SNSGHSSFWVDFDDTENKSVLRGGPLTSYRQLQVHLSWGSADDHGSHIIVDGVSYAAELH 132

Qy 121 VWHNNSDORYPSFVEAAHEDPGGLAVLGFLQIGEPNSQOKITDTLDLKEXKGKQTRETNF 180

Db 133 VWHNNSDORYPSFVEAAHEDPGGLAVLGFLQIGEPNSQOKITDTLDLKEXKGKQTRETNF 192

Qy 181 DLISLIPPSWDWYTYPGSLTVPPLESTWVILQOPINISSQOLAKERSLCTAEGRAAA 240

Db 193 DLISLIPPSWDWYTYPGSLTVPPLESTWVILQOPINISSQOLAKERSLCTAEGRAAA 252

Qy 241 FL 242

Db 253 PL 254

RESULT 4

ID ABJ37885 standard; protein: 262 AA.

XX ABJ37885;

XX 22-MAY-2003 (first entry)

NOVX protein sequence SEQ ID No 16.

XX Hepatotoxic; immunosuppressive; cardiotonic; hypertensive; tranquilizer;
 KW vulneraric; virucide; antibacterial; protozoacide; fungicide; nootropic;
 KW antiparasitic; neuroprotective; cerebroprotective; antiparkinsonian;
 KW antiaddictive; analgesic; dermatological; keratolytic;
 KW antiseborrheic; antiarthritic; antiinflammatory; anti-HIV;

KW cyrostatic; antiasthmatic; antipruritic; hypotensive; osteopathic;

KW antiulcer; anorectic; antidiabetic; antiallergic; haemostatic;

KW neuroleptic; antidepressant; antinfertility; NOVX; human disease;

KW NOVX-associated disorder; trauma; viral; bacterial; fungal; protozoal;

KW parasitic infection; Alzheimer's disease; stroke; forensic biology;

KW immunogen; non-human transgenic animal; gene therapy.

XX Unidentified.

PN WO2002B1517-A2.

XX PD 17-OCT-2002.

XX PP 22-JAN-2002; 2002WO-US002064.

XX PR 19-JAN-2001; 2001US-0262892P.
 PR 23-JAN-2001; 2001US-0262892P.
 PR 24-JAN-2001; 2001US-0263799P.
 PR 25-JAN-2001; 2001US-0264117P.
 PR 26-JAN-2001; 2001US-0264139P.
 PR 30-JAN-2001; 2001US-0264478P.
 PR 02-MAR-2001; 2001US-026351P.
 PR 14-MAR-2001; 2001US-0275927P.
 PR 14-MAR-2001; 2001US-0275919P.
 PR 15-MAR-2001; 2001US-0274494P.
 PR 20-MAR-2001; 2001US-027358P.

PR 23-MAR-2001; 2001US-027B151P.
 PR 29-MAR-2001; 2001US-0279817P.
 PR 20-APR-2001; 2001US-0285140P.
 PR 20-APR-2001; 2001US-0285141P.
 PR 30-APR-2001; 2001US-0287848P.
 PR 17-MAY-2001; 2001US-0291701P.
 PR 08-JUN-2001; 2001US-0296900P.
 PR 10-JUL-2001; 2001US-0304333P.
 PR 10-JUL-2001; 2001US-0304355P.
 PR 12-JUL-2001; 2001US-0304866P.
 PR 09-AUG-2001; 2001US-0311299P.
 PR 13-AUG-2001; 2001US-0311975P.
 PR 16-AUG-2001; 2001US-0312931P.
 PR 18-OCT-2001; 2001US-0330227P.
 PR 29-NOV-2001; 2001US-0334198P.
 XX (CURAGEN CORP.
 PA XX Decristofaro MF, Padigaru M, Miller C, Tchernev V, Zhong H; Gerlach R, Rastelli L;
 PI Khong M, Anderson D, Ballinger R, Gerlach R, Paituraian M;
 PI Kekuda R, Guo X, Zerhusen B, Andrew D, Mezes J, Shambets RA, Gusev V;
 PI Burgess CE, Eisen A, Woltenc A, Baumgartner J, Verner CAM, Taupier RJ, Pena C, Shenoy S, Li L, Caizman S, Boldog F;
 PI Fernandes E, Smithson G, Malyankar U, Tailion B, Liu X;
 XX WPI; 2003-058504/05.
 DR N PSSDB; ABT33350.
 XX New polypeptides, designated as NOVX, useful for diagnosing and treating
 PT infections, neurological diseases, cancer, allergy, and bone,
 PT immunological, skin, renal, brain, muscle and autoimmune disorders.
 XX Claim 1; Page 66: 672PP; English.
 XX The invention relates to a novel isolated polypeptide, designated NOVX
 CC (NOV1 - 33), consisting of a mature form of one of 61 sequences, given in
 CC the specification, or its variant, where amino acid residue(s) in the
 CC variant differ from the mature form, provided that the variant differs in
 CC not more than 15 % of the amino acids from the sequence of the mature
 CC form. The NOVX polypeptides, nucleic acids encoding the polypeptides, and
 CC an antibody to the polypeptides, are useful for treating or preventing a
 CC NOVX-associated disorder in humans and for treating a syndrome associated
 CC with human disease (NOVX-associated disorder). NOVX polypeptides and
 CC the encoding nucleic acids, are useful for determining the presence of or
 CC predisposition to a disease associated with altered levels of NOVX
 CC polypeptide and polynucleotide, by measuring the level of polypeptide
 CC expression or the amount of nucleic acid from a mammal and comparing it
 CC with another mammal not having or not predisposed to the disease. NOVX
 CC polypeptide is also useful for identifying an agent that binds to NOVX
 CC and a cell expressing NOVX is useful for identifying an agent that
 CC modulates the expression or activity of NOVX. The antibodies and a
 CC polypeptide having 95 % sequence identity to NOVX are useful
 CC for treating a pathological state in a mammal. The antibodies are also
 CC useful for determining the presence or amount of NOVX in a sample. NOVX
 CC polypeptides, polynucleotides and antibodies specific for the
 CC including trauma, viral, bacterial, fungal, protozoal, and parasitic
 CC infections. They can also treat disorders such as e.g., Alzheimer's
 CC disease or a stroke. The NOVX encoding nucleic acids are useful for
 CC expressing the NOVX proteins, to detect NOVX mRNA, or a genetic lesion in
 CC a NOVX gene and to modulate NOVX activity. NOVX sequences are also useful
 CC for identifying a cell or tissue type in a biological sample, to amplify
 CC DNA sequences from very small biological samples such as tissues e.g.
 CC hair or skin or body fluids in forensic biology and as primers and probes
 CC for use in identifying and/or cloning NOVX homologues in other cell
 CC types. The NOVX proteins are useful as an immunogen to generate
 CC antibodies which are useful for diagnostically monitoring protein levels
 CC and modulating NOVX activity. Cells comprising NOVX nucleic acids are
 CC useful for producing non-human transgenic animals which are useful for
 CC studying the function and/or activity of NOVX protein and for identifying
 CC and/or evaluating modulators of NOVX protein activity. The NOVX nucleic
 CC acids can be used in gene therapy. This sequence represents a NOVX
 CC protein of the invention

KW cardiovascular disorder; cardiac arrest; cerebrovascular disorder;
 cerebral ischaemia; angiogenesis; nervous system disorder; infection;
 Alzheimer's disease; ocular disorder; corneal infection; wound healing;
 XX skin aging; food additive; preservative.

OS Homo sapiens.
 PN WO200001748-A1.
 XX PD 19-OCT-2000.
 XX DF 06-APR-2000; 2000WO-US008982.
 XX PR 09-APR-1999; 99US-0128695P.
 PR 14-JAN-2000; 2000US-0176063P.
 XX PA (HUMA-) HUMAN GENOME SCI INC.
 XX PT Rosen CA, Ruben SM, Komatsoulis G;
 XX DR WPI; 2000-638566/61.
 XX PT New nucleic acid molecules encoding 48 human secreted proteins for
 diagnosis, prevention, treating or ameliorating medical conditions and
 PT used as food additives or preservatives.
 XX Disclosure: Page 458-459; 480pp; English.
 XX AAF22316 to AAF22363 encode the human secreted proteins given in AAB63049
 CC to AAB63096. AAB63132 represent more human secreted proteins
 CC and polypeptides homologous to them. Human secreted proteins have
 CC activities based on the tissues and cells the genes are expressed in.
 CC Examples of activities include: immunosuppressive; antiarthritic;
 CC antirheumatic; antiproliferative; cytostatic; cardiotonic; vasotropic;
 CC cerebroprotective; nootropic; neuroprotective; antibacterial; virucide;
 CC fungicide; ophthalmological; and vulnerary. The polynucleotides and
 CC proteins can be used to prevent, treat or ameliorate a medical condition
 CC in e.g. humans, mice, rabbits, goats, horses, cats, dogs, chickens or
 CC sheep. They are also used in diagnosing pathological condition or
 CC susceptibility to a pathological condition. Disorders which are diagnosed
 CC or treated include autoimmune diseases e.g. rheumatoid arthritis,
 CC neoplasms of the breast or liver, cardiovascular disorders e.g.
 CC cerebral ischaemia, angiogenesis, nervous system disorders e.g.
 CC Alzheimer's disease, infections caused by bacteria, viruses and fungi and
 CC ocular disorders e.g. corneal infection. The polypeptides can also be
 CC used to aid wound healing and epithelial cell proliferation, to prevent
 CC skin aging due to sunburn, to maintain organs before transplantation, for
 CC supporting cell culture of primary tissues, to regenerate tissues and in
 CC chemotaxis. The polypeptides can also be used as a food additive or
 CC preservative to increase or decrease storage capabilities, fat content,
 CC lipid, protein, carbohydrate, vitamins, minerals, cofactors and other
 CC nutritional components. AAF22307 to AAF22315 and AAB63048 represent
 CC sequences used in the exemplification of the present invention.
 XX Sequence 184 AA;

Query Match 67.0%; Score 864; DB 3; Length 184;
 Best Local Similarity 98.8%; Pred. No. 1e-81; Mismatches 1; Indels 0; Gaps 0;
 Matches 165; Conservative 165; Mismatches 1; Indels 0; Gaps 0;

QY 78 SVLRRGPAITGFLQIQLQPSYRQRQVHLWGSADDHGSSRHIVGVSYAAELHVHWNSDKYPFVEAAH 137
 1 NVLRQGPFLQIQLQPSYRQRQVHLWGSADDHGSSRHIVGVSYAAELHVHWNSDKYPFVEAAH 60
 DB 138 EPDGIAVLGVFLQIQLQPSYRQRQVHLWGSADDHGSSRHIVGVSYAAELHVHWNSDKYPFVEAAH 197
 61 EPDGIAVLGVFLQIQLQPSYRQRQVHLWGSADDHGSSRHIVGVSYAAELHVHWNSDKYPFVEAAH 120
 DB 121 VHVHNSDKYPFVEAAHEPDGLAVLGFLQIQLQPSYRQRQVHLWGSADDHGSSRHIVGVSYAAELHVHWNSDKYPFVEAAH 120
 DB 121 VAHNSAKSSLBAAKSKADGLAVLGFLQIQLQPSYRQRQVHLWGSADDHGSSRHIVGVSYAAELHVHWNSDKYPFVEAAH 180
 QY 198 SITVPPLIESVTWIVLKVQPINISSQLAKERSLSLTCAESEAAAFL 242
 DB 121 SITVPPLIESVTWIVLKVQPINISSQLAKERSLSLTCAESEAAAFL 165
 DB 181 DPSTLIPSSDFTWVTPGSLTHPPYESVTWICKESESISVSEQAQFRSLLSNVEGDN 239

Query Match 62.4%; Score 804; DB 5; Length 261;
 Best Local Similarity 59.4%; Pred. No. 3.3e-75; Mismatches 53; Indels 0; Gaps 0;
 Matches 142; Conservative 142; Mismatches 53; Indels 0; Gaps 0;

QY 1 MSRSWSWGRREHNGIHWWKEFFP1ADGDQSPIEIKTKEVYKDSSIRPLSIKYDSSAKLI 60
 1 :|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:
 1 MASPDWGYDDKGDEQWSKLYPLANGNNQSPVDLKTSETKHDTSKPSVSYNTATAKEI 60
 DB 61 SNSGHSHFSNYDFFDTENPKSVLRRGGPLTGSYRLQWHLWGSADDHGSSRHIVGVSYAAELHVHWNSDKYPFVEAAH 120
 DB 61 INVGHSFHYNEDDNRNSVLUKGGSFDSYRLQFQFHEHGSTNEHGESENTHDVGYSYAAELHVHWNSDKYPFVEAAH 120
 QY 121 VHVHNSDKYPFVEAAHEPDGLAVLGFLQIQLQPSYRQRQVHLWGSADDHGSSRHIVGVSYAAELHVHWNSDKYPFVEAAH 180
 DB 121 VAHNSAKSSLBAAKSKADGLAVLGFLQIQLQPSYRQRQVHLWGSADDHGSSRHIVGVSYAAELHVHWNSDKYPFVEAAH 180

| | | | |
|----------|---|--|--|
| RESULT 8 | | RESULT 9 | |
| ADA10962 | ID ADA10962 standard; protein; 261 AA. | AAG73863 standard; protein; 263 AA. | |
| XX | AC ADA10962; | AC AAG73863; | |
| XX | DT 06-NOV-2003 (first entry) | DT 03-SEP-2001 (first entry) | |
| XX | DE Human cDNA differentially expressed in colon cancer #54 product. | DE Human colon cancer antigen protein SEQ ID NO:4627. | |
| XX | DB Human cDNA differentially expressed in colon cancer #54 product. | XX Human; colon cancer; colon cancer antigen; diagnosis; detection; | |
| XX | KW differential expression; colon cancer; cancer; human. | KW colorectal carcinoma; chromosome 8. | |
| XX | OS Homo sapiens. | XX Homo sapiens. | |
| XX | PN US2002160382-A1. | PN WO200122920-A2. | |
| XX | PD 31-OCT-2002. | PD 05-APR-2001. | |
| XX | PF 11-OCT-2001; 2001US-00981353. | PF 28-SEP-2000; 2000WO-US026524. | |
| XX | PR 11-OCT-2000; 2000US-0239841P. | PR 29-SEP-1999; 99US-0157137P. | |
| XX | PA (LASEK A W. (JONE/) JONES D A. | PR 03-NOV-1999; 99US-0163280P. | |
| XX | PI Lasek AW, Jones DA; | PI Ruben SM, Barash SC, Birse CE, Rosen CA; | |
| XX | DR 2003-265756/26. | XX WPI; 2001-235357/24. | |
| XX | DR N-PSDB; ADA10961. | DR N-PSDB; AAH33294. | |
| XX | PT New combination comprising cDNAs that are differentially expressed in colon disorder, useful for diagnosing, treating, staging or monitoring treatment for colon cancers. | XX Nucleic acids encoding 4277 human colon cancer-associated polypeptides, useful for preventing, diagnosing and/or treating colorectal cancers. | |
| XX | PT Example 14; SEQ ID NO 80; 231pp; English. | XX Claim 11; Page 6427-6429; 9803pp; English. | |
| XX | PT The invention relates to a combination comprising cDNAs that are differentially expressed in colon disorder. The methods and compositions of the present invention are useful for diagnosing, treating, staging or monitoring treatment for colon cancer. They are also useful in high throughput methods for using cDNAs to detect differential expression of nucleic acids in a sample, screening molecules or compounds to identify a ligand which specifically binds a cDNA and using a protein to screen molecules or compounds to identify at least one ligand which specifically binds the protein. The present sequence represents the amino acid sequence of a human cDNA differentially expressed in colon cancer protein. | XX AAH32943 to AAH37195 and AAG77788 represent human colon cancer-associated nucleic acid molecules (N) and proteins (P), where the proteins are collectively called colon cancer antigens. The colon cancer antigens have cytostatic activity and can be used in gene therapy and vaccine production. N and P may be used in the prevention, diagnosis and treatment of diseases associated with inappropriate P expression. For example, N and P may be used to treat disorders associated with decreased expression by rectifying mutations or deletions in a patient's genome that affect the activity of P by expressing inactive proteins or to supplement the patients own production of P. Additionally, N may be used to produce the colon cancer-associated Ps, by inserting the nucleic acids into a host cell and culturing the cell to express the proteins. N and P can be used in the prevention, diagnosis and treatment of colorectal carcinomas and cancers. AAH37196 to AAH37204 and AAB77789 represent sequences used in the exemplification of the present invention. N.B. Pages 66 to 682 and page 7053 of the sequence listing were missing at time of publication, meaning no sequences are present for SEQ ID NO:1027 to 1052, 7921 and 7922. | |
| XX | PS Sequence 261 AA; | XX Sequence 263 AA; | |
| XX | Score 62.4%; Score 804; DB 6; Length 261; | Query Match 62.4%; Score 804; DB 4; Length 263; | |
| XX | Best Local Similarity 59.4%; Pred. No. 3.3e-75; | Best Local Similarity 59.4%; Pred. No. 3.3e-75; | |
| XX | Matches 142; Conservative 44; Mismatches 53; Indels 0; Gaps 0; | Matches 142; Conservative 44; Mismatches 53; Indels 0; Gaps 0; | |
| SQ | 1 MSRLSWGYREHNGPITHKEFFPIADGDDQOSPIEKTKEVKYDSSLRPLSIKYDSSAKII 60 | QY 1 MSRLSWGYREHNGPITHKEFFPIADGDDQOSPIEKTKEVKYDSSLRPLSIKYDSSAKII 60 | |
| Db | 1 MASPDGMDYDDKNGPDDNPKLPIANGNNSPVDIKTSETKHDTSKLPISYNPATAKEI 60 | Db 3 MASPDGMDYDDKNGPDDNPKLPIANGNNSPVDIKTSETKHDTSKLPISYNPATAKEI 62 | |
| Qy | 61 SNSGHSFNVDDDTENKSVLNGPLTGSYRLQVHLWGADDKHSSEHTYDGVSZAELH 120 | Qy 61 SNSGHSFNVDDDTENKSVLNGPLTGSYRLQVHLWGADDKHSSEHTYDGVSZAELH 120 | |
| Db | 61 INVGHSHSNFDNNPKLPIANGNNSPVDIKTSETKHDTSKLPISYNPATAKEI 60 | Db 61 INVGHSHSNFDNNPKLPIANGNNSPVDIKTSETKHDTSKLPISYNPATAKEI 60 | |
| Qy | 61 INVGHSHSNFDNNPKLPIANGNNSPVDIKTSETKHDTSKLPISYNPATAKEI 60 | Qy 61 INVGHSHSNFDNNPKLPIANGNNSPVDIKTSETKHDTSKLPISYNPATAKEI 62 | |
| Db | 61 INVGHSHSNFDNNPKLPIANGNNSPVDIKTSETKHDTSKLPISYNPATAKEI 60 | Db 63 INVGHSHSNFDNNPKLPIANGNNSPVDIKTSETKHDTSKLPISYNPATAKEI 120 | |
| Qy | 121 VVHWNSDKYPSFVEAAHEPDGLAVLGFLQIGEPNSQLOXITDTLDISIKEKGKQTFTNF 180 | Db 121 VVHWNSDKYPSFVEAAHEPDGLAVLGFLQIGEPNSQLOXITDTLDISIKEKGKQTFTNF 180 | |
| Db | 121 VAHWNSAKYSSLAEEASKADGLAVGTVLMKVGEANPKLQVWLDALOAIKYGKRAPFTNF 180 | Db 121 VAHWNSAKYSSLAEEASKADGLAVGTVLMKVGEANPKLQVWLDALOAIKYGKRAPFTNF 180 | |
| Qy | 181 DLLSLPPSWDYWTYPGSLTYPPLLESVTWIKQPINISQQLAKERSLICLTAGEAA 239 | Qy 181 DLLSLPPSWDYWTYPGSLTYPPLLESVTWIKQPINISQQLAKERSLICLTAGEAA 239 | |
| Db | 181 DPSTLPPSWDYWTYPGSLTYPPLLESVTWIKQPINISQQLAKERSLICLTAGEAA 239 | Db 181 DPSTLPPSWDYWTYPGSLTYPPLLESVTWIKQPINISQQLAKERSLICLTAGEAA 239 | |
| 123 | 123 VAHWNSAKYSSLAEEASKADGLAVGTVLMKVGEANPKLQVWLDALOAIKYGKRAPFTNF 180 | 123 VAHWNSAKYSSLAEEASKADGLAVGTVLMKVGEANPKLQVWLDALOAIKYGKRAPFTNF 180 | |

| | | | | | |
|-----------------------|--|---|--|---|---|
| QY | 181 DLLSLIPPSWDYWNTPGSLTVPPPLLESVTWVLKOPINISSQLAKFRSLCTAREAA 182 : : : : : : : : : : : : : 183 DESTLILPSLDENTWNTYPSSLTTHPLYESVTWICKESTISVSSEQLAQFRLLSNVEGDNA | 239 241 | Qy | 66 SENVDDDTENKSVLRGGPLTGSYRLRQVHLHWSADDHGSEHTIVDGYSYAAELHVYHWN 125 67 SPHNVEDNDNRSVLRKGPFEDDSYRLRQVHLHWSADDHGSEHTIVDGYSYAAELHVYHWN 124 | |
| RESULT 10 | AAB59588 | standard; protein; 260 AA. | Db | 126 SDKYPSEFVEAAHEPDGLAVLGFLQIGEPNSQLOKITTDLSIKEKGKQTRFTNFDLSSL 185 | |
| ID | AAB59588 | | Db | 125 SARYSSIAEARSQADGLAVIGULMKGEANPKLQKVLDALQIKTGKRAFTNFDPSTL 184 | |
| XX | AC | AAB59588; | Qy | 186 LPSSWDWTYFGSLTPPLLESVTWVLKOPINISSQLAKFRSLCTAEGEEA 239 | |
| XX | AC | | Db | 185 LPSSLDWTYFGSLTPPLLESVTWICKESTISVSSEQLAQFRLLSNVEGDNA 238 | |
| DT | 28-MAR-2001 | (first entry) | | | |
| XX | DE | Human carbonic anhydrase isoform #1. | | | |
| XX | KW | protein tyrosine phosphatase; PTase; vulnerability; cytostatic; | | | |
| XX | KW | antiinflammatory; antidiabetic; viral infection; inflammation; cancer; | | | |
| XX | KW | receptor-type protein tyrosine phosphatase beta; RPTPbeta; | | | |
| XX | KW | carbonic anhydrase; CAH; diabetes mellitus. | | | |
| XX | OS | Homo sapiens. | | | |
| XX | PP | 9JUS-00081949. | | | |
| PN | US6160090-A. | | | | |
| XX | PR | 90US-00551270. | | | |
| PD | 12-DEC-2000. | | | | |
| XX | PR | 91US-00654188. | | | |
| PP | 23-JUN-1993; | | | | |
| XX | PR | 92US-00961235. | | | |
| PP | 11-JUL-1990; | | | | |
| XX | PR | 92US-00015973. | | | |
| PD | 15-OCT-1992; | | | | |
| XX | PR | 93US-00015973. | | | |
| PP | 10-FEB-1993; | | | | |
| XX | PA | (UYNY) UNIV NEW YORK STATE. | | | |
| XX | PA | Margolis RU, Grunet MH, Barnea G, Schlessinger J; | | | |
| PI | XX | DR | WPI; 2001-070117/08. | | |
| XX | PT | Novel receptor type protein tyrosine phosphatase beta protein used to modulate normal cellular processes of differentiation, metabolism, cell cycle by competing with endogenous transmembrane receptors for ligands. | | | |
| PT | PT | The present sequence is given in a specification relating to a novel receptor type protein tyrosine phosphatase beta (RPTPbeta) protein or glycoprotein. The receptor and its ligands are useful for developing compounds and strategies for modifying cellular processes e.g. normal cellular processes such as differentiation, metabolism, cell cycle control, wound healing and neuronal function, cellular behaviour such as mobility, migration, and contact inhibition, in addition to abnormal or potentially deleterious processes such as virus-receptor interactions, inflammation, cellular transformation to a cancerous state, and the development of Type 2, insulin independent, diabetes mellitus, under the control of the receptor protein tyrosine phosphatases. The receptors or their ligands may be used directly to modulate processes such as those mentioned above. They act to compete with endogenous transmembrane receptor molecules for available RPTPases. RPTPases and/or their ligands, may also be used to screen for additional molecules that can act to modulate the activity of cellular processes | | | |
| PS | PS | Example: Fig 7; 45pp; English. | | | |
| XX | CC | The present sequence is given in a specification relating to a novel receptor type protein tyrosine phosphatase beta (RPTPbeta) protein or glycoprotein. The receptor and its ligands are useful for developing cellular processes such as differentiation, metabolism, cell cycle control, wound healing and neuronal function, cellular behaviour such as mobility, migration, and contact inhibition, in addition to abnormal or potentially deleterious processes such as virus-receptor interactions, inflammation, cellular transformation to a cancerous state, and the development of Type 2, insulin independent, diabetes mellitus, under the control of the receptor protein tyrosine phosphatases. The receptors or their ligands may be used directly to modulate processes such as those mentioned above. They act to compete with endogenous transmembrane receptor molecules for available RPTPases. RPTPases and/or their ligands, may also be used to screen for additional molecules that can act to modulate the activity of cellular processes | | | |
| XX | CC | Sequence 260 AA; | | | |
| SQ | Query Match | 62.2%; Score 802; DB 4; Length 260; | Qy | 61.4%; Score 791.5; DB 2; Length 260; | |
| Best Local Similarity | 60.3%; | Pred. No. 5.3e-75; | Db | 62.1%; Pred. No. 6.6e-74; | |
| Matches 141; | Conservative | Mismatches 43; | Mismatches 55; | Indels 1; Gaps 1; | |
| Qy | 6 WGYREHNGPITHWKKEFFPTADGDOQSPIEKTKTKEYTDSSLRPLSISKYDPSAKLISNSGH 65 5 WGXRKHNGPHEWHDFPIANGERSPVDIDTHAKYDPSLKPVLPSYDATSLSLELNCH 64 | Db | 66 SPNDTFDDDTENKSVLRGGLTGSYRLRQVHLHWSADDHGSEHTIVDGYSYAAELHVYHWN 125 65 AFNTEFDSDQDKAVLKGGPLDGTURLIQPFHMSLQDGQGSEHTIVDGYSYAAELHVYHWN 124 | Qy | 126 SDKYPSEFVEAAHEPDGLAVLGFLQIGEPNSQLOKITTDLSIKEKGKQTRFTNFDLSSL 185 |
| Qy | 5 WGYYDDKNGPQEWQSKLYPIANGNNQSPVIDKTSKLPISVSYNPATAKEIINVGH 64 | Db | | | |

| | | | | | |
|----|--|-----------------|----|---|-----|
| Db | 125 T-KYGFQKAVQEPDGLAVLGFQKLGKQVVDLSIKPGKSAADFTNIDPRGL | 183 | Db | 4 WGYGRKINGPEWHKDPIAKERQSPVDDIDHTAKYDPSSLPLSVSYDQATSLRLNNNGH | 63 |
| Qy | 186 LPPSNWDYTYPGSLTVPPLIESVTWIVLKQPNINSSCOLAKERSLIICTAEGE | 237 | Qy | 66 SENVDFDDTENKSVLRCGPGLTGSYRLRQVHLHWGSADDHGSEHIVDGVSYAAELHYHWN | 125 |
| Db | 184 LPESLDWYTYPGSLTVPPLIESVTWIVLKQPNINSSCOLAKERSLIICTAEGE | 235 | Db | 64 AFNEFEDSQDAVLKGPGPLTYRIJQFHPHGGSDGQGSHTVDRKKYIAELHVWN | 123 |
| | AAB59589 | | | | |
| | ID AAB59589 standard; protein; 259 AA. | | | | |
| | XX | | | | |
| | AAB59589; | | | | |
| | XX | | | | |
| | 28-MAR-2001 (first entry) | | | | |
| | XX | | | | |
| | Human carbonic anhydrase isoform #2. | | | | |
| | XX | | | | |
| | KW Human; protein tyrosine phosphatase; PTase; vulnerability; cytostatic; antiinflammatory; antidiabetic; viral infection; inflammation; cancer; receptor-type protein tyrosine phosphatase beta; RPTPbeta; carbonic anhydrase; CAH; diabetes mellitus. | | | | |
| | XX | | | | |
| | Homo sapiens. | | | | |
| | XX | | | | |
| | OS US6160090-A. | | | | |
| | PN | | | | |
| | KW XX | | | | |
| | PD 12-DEC-2000. | | | | |
| | XX | | | | |
| | PF 23-JUN-1993; | 93US-000981929. | | | |
| | XX | | | | |
| | PR 11-JUL-1990; | 90US-00551270. | | | |
| | PR 26-FEB-1991; | 91US-00654118. | | | |
| | PR 15-OCT-1992; | 92US-00961235. | | | |
| | PR 10-FEB-1993; | 93US-00015973. | | | |
| | XX | | | | |
| | PA (UINY) UNIV NEW YORK STATE. | | | | |
| | XX | | | | |
| | Margolis RU, Grumet MH, Barnea G, Schlessinger J; | | | | |
| | XX | | | | |
| | WPI; 2001-070117/08. | | | | |
| | XX | | | | |
| | PT Novel receptor type protein tyrosine phosphatase beta protein used to modulate normal cellular processes of differentiation, metabolism, cell cycle by competing with endogenous transmembrane receptors For ligands. | | | | |
| | XX | | | | |
| | PS Example: Fig 7; 45pp; English. | | | | |
| | XX | | | | |
| | CC The present sequence is given in a specification relating to a novel receptor-type protein tyrosine phosphatase beta (RPTPbeta) protein or glycoprotein. The receptor and its ligands are useful for developing compounds and strategies for modifying cellular processes e.g. normal cellular processes such as differentiation, metabolism, cell cycle control, wound healing and neuronal function. Cellular behaviour such as motility, migration, and contact inhibition, in addition to abnormal or inflammatory, deleterious processes such as virus-receptor interactions, cellular transformation to cancerous state, and the development of Type 2, insulin independent, diabetes mellitus, under the control of the receptor protein tyrosine phosphatases. The receptors or their ligands may be used directly to modulate processes such as those mentioned above. They act to compete with endogenous transmembrane receptor molecules for available ligands, thus reducing or inhibiting ligand binding to endogenous RPTPases. RPTPases and/or their ligands, may also be used to screen for additional molecules that can act to modulate the activity of cellular processes | | | | |
| | CC SQ Sequence 259 AA; | | | | |
| | CC The present sequence is given in a specification relating to a novel receptor-type protein tyrosine phosphatase beta (RPTPbeta) protein or glycoprotein. The receptor and its ligands are useful for developing compounds and strategies for modifying cellular processes e.g. normal cellular processes such as differentiation, metabolism, cell cycle control, wound healing and neuronal function. Cellular behaviour such as motility, migration, and contact inhibition, in addition to abnormal or inflammatory, deleterious processes such as virus-receptor interactions, cellular transformation to cancerous state, and the development of Type 2, insulin independent, diabetes mellitus, under the control of the receptor protein tyrosine phosphatases. The receptors or their ligands may be used directly to modulate processes such as those mentioned above. They act to compete with endogenous transmembrane receptor molecules for available ligands, thus reducing or inhibiting ligand binding to endogenous RPTPases. RPTPases and/or their ligands, may also be used to screen for additional molecules that can act to modulate the activity of cellular processes | | | | |
| | CC XX | | | | |
| | CC Query Match 61.2%; Score 788.5; DB 4; Length 259; | | | | |
| | CC Best Local Similarity 61.6%; Pred. No. 1.3e-73; | | | | |
| | CC Matches 143; Conservative 33; Mismatches 55; Indels 1; Gaps 1; | | | | |
| | CC Qy 6 WGYREINGPIHWKEFFBIAGDQQSIEKVKYDSSAKILISNSGH 65 | | | | |

polypeptides or their antibodies. The polynucleotide or the compound that modulates its activity is useful for preparing a medicament for treating pain (e.g. spinal segmental nerve injury (Chung), chronic constriction injury (CCI) and spared nerve injury (SNI)) in an animal (e.g. gene therapy). The sequence presented is a human protein (shown in Table 2 of the specification) which is differentially expressed during pain. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic form directly from WIPO at ftp://wipo.int/pub/published_pct_sequences.

Sequence 259 AA:

| | | | | |
|---------------------------|-------|---|-----------|-------------|
| Query Match | 61.2% | Score 788.5; | DB 7; | Length 259; |
| Best Local Similarity | 61.6% | Pred. No. 1.3e-73; | | |
| Matches 143; Conservative | 33; | Mismatches 55; | Indels 1; | Gaps 1; |
| Qy | 6 | WGYYREENGPIHNNKEEPIADGQDQSPLEIKTKEVKYDSLRPLSIKYDPSSAKIISNSGH 65 | | |
| Db | 4 | WGYYGKHNHPEAHWIKDFPAAKGERQSPVIDTHAKYDLSKPUVSQATSLRILNGH 63 | | |
| Qy | 66 | SFNPVDFDTENKSVLRGPAPLTSYRKLQHHLNGSADHGSEHIVDGYSYAAELHVWN 125 | | |
| Db | 64 | AENVEFDSDQDKAVLKGGPGLDTYRLQHFWHGSLDQGSEHTVDKCKYAAELHVWN 123 | | |
| Qy | 126 | SDKYPKFEEAAAHBDGFLAVLGVLQLQGPNSQLQKITLTDSTKEKGHQTRFTNFDLISL 185 | | |
| Db | 124 | T-KYGDGRFAVQCDPGLFLVKGAKPGIQKVVWLDLRTKGQSADFTNEDPFGI 182 | | |
| Qy | 186 | LPPSWDNTYTPGSLTVPPILESYTWIVLQKOPINISSQOLAKERFSLLCTAEGE 237 | | |
| Db | 183 | LPPESLDNTYTPGSLTTTPPILECYTWIVLQKEPISVSSBVLKFRKLNFNGEGE 234 | | |

RESULT 1.4

ID ADE63735
XX ADE63735; DT 29-JAN-2004 (first entry)
XX Human Protein P00918, SEQ ID NO 9679.

Human; pain; neuronal tissue; gene therapy;
spinal segmental nerve injury; chronic constriction injury; CCI;
spared nerve injury; SNI; Chung.
XX Homo sapiens.
XX WO2003016475-A2.

PI Woolf C, D'urso D, Befort K, Costigan M;

XX PD 27-FEB-2003.

XX PF 14-AUG-2002; 2002WO-US025765.

XX PR 14-AUG-2001; 2001US-0312147P.
PR 01-NOV-2001; 2001US-0346382P.
PR 26-NOV-2001; 2001US-0333347P.

XX PA (GEHO) GEN HOSPITAL CORP.
PA (FARB) BAYER AG.

XX PN WO2003016475-A2.

XX DR 2003-02812/26.

XX GENBANK; P00918.

XX New composition comprising two or more isolated polypeptides, useful for preparing a medicament for treating pain in an animal.
XX PS Claim 1; Page; 1017pp; English.
XX CC The invention discloses a composition comprising two or more isolated rat

or human polynucleotides or a polynucleotide which represents a fragment, a derivative or allelic variation of the nucleic acid sequence. Also claimed are a vector comprising the novel polynucleotide, a host cell comprising the vector, a method for identifying a nucleotide sequence which is differentially regulated in an animal subjected to pain and a kit to perform the method, an array, a method for identifying an agent that increases or decreases the expression of the polynucleotide sequence that is differentially expressed in neuronal tissue of a first animal subjected to pain, a method for identifying a compound which regulates the expression of a polynucleotide sequence which is differentially expressed in an animal subjected to pain, a method for identifying a compound that regulates the activity of one or more of the polypeptides given in the pharmaceutical composition, a method for producing a pharmaceutical composition, a method for identifying a compound or small molecule that regulates the activity in an animal of one or more of the polypeptides given in the specification, a method for identifying a compound useful in treating pain and a pharmaceutical composition comprising the one or more polypeptides or their antibodies. The polynucleotide modulates its activity for preparing a medicament for treating pain, chronic constriction injury (CCI) and spared nerve injury (SNI) in an animal (e.g. gene therapy). The sequence presented is a human protein (shown in Table 2 of the specification) which is differentially expressed during pain. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic form directly from WIPO at ftp://wipo.int/pub/published_pct_sequences.

Sequence 259 AA;

| | | | | |
|-------|-------------------------------|---|-----------|-------------|
| Query | Match 61.2% | Score 788.5; | DB 7; | Length 259; |
| | Best Local Similarity 61.6%; | Pred. No. 1.3e-73; | | |
| | Matches 143; Conservative 33; | Mismatches 55; | Indels 1; | Gaps 1; |
| Qy | 6 | WGYYREENGPIHNNKEEPIADGQDQSPLEIKTKEVKYDSLRPLSIKYDPSSAKIISNSGH 65 | | |
| Db | 4 | WGYYGKHNHPEAHWIKDFPAAKGERQSPVIDTHAKYDLSKPUVSQATSLRILNGH 63 | | |
| Qy | 66 | SFNPVDFDTENKSVLRGPAPLTSYRKLQHHLNGSADHGSEHIVDGYSYAAELHVWN 125 | | |
| Db | 64 | AENVEFDSDQDKAVLKGGPGLDTYRLQHFWHGSLDQGSEHTVDKCKYAAELHVWN 123 | | |
| Qy | 126 | SDKYPKFEEAAAHBDGFLAVLGVLQLQGPNSQLQKITLTDSTKEKGHQTRFTNFDLISL 185 | | |
| Db | 124 | T-KYGDGRFAVQCDPGLFLVKGAKPGIQKVVWLDLRTKGQSADFTNEDPFGI 182 | | |
| Qy | 186 | LPPSWDNTYTPGSLTVPPILESYTWIVLQKOPINISSQOLAKERFSLLCTAEGE 237 | | |
| Db | 183 | LPPESLDNTYTPGSLTTTPPILECYTWIVLQKEPISVSSBVLKFRKLNFNGEGE 234 | | |

Sequence 259 AA;

| | | | | |
|-------|-------------------------------|---|-----------|-------------|
| Query | Match 61.2% | Score 788.5; | DB 7; | Length 259; |
| | Best Local Similarity 61.6%; | Pred. No. 1.3e-73; | | |
| | Matches 143; Conservative 33; | Mismatches 55; | Indels 1; | Gaps 1; |
| Qy | 6 | WGYYREENGPIHNNKEEPIADGQDQSPLEIKTKEVKYDSLRPLSIKYDPSSAKIISNSGH 65 | | |
| Db | 4 | WGYYGKHNHPEAHWIKDFPAAKGERQSPVIDTHAKYDLSKPUVSQATSLRILNGH 63 | | |
| Qy | 66 | SFNPVDFDTENKSVLRGPAPLTSYRKLQHHLNGSADHGSEHIVDGYSYAAELHVWN 125 | | |
| Db | 64 | AENVEFDSDQDKAVLKGGPGLDTYRLQHFWHGSLDQGSEHTVDKCKYAAELHVWN 123 | | |
| Qy | 126 | SDKYPKFEEAAAHBDGFLAVLGVLQLQGPNSQLQKITLTDSTKEKGHQTRFTNFDLISL 185 | | |
| Db | 124 | T-KYGDGRFAVQCDPGLFLVKGAKPGIQKVVWLDLRTKGQSADFTNEDPFGI 182 | | |
| Qy | 186 | LPPSWDNTYTPGSLTVPPILESYTWIVLQKOPINISSQOLAKERFSLLCTAEGE 237 | | |
| Db | 183 | LPPESLDNTYTPGSLTTTPPILECYTWIVLQKEPISVSSBVLKFRKLNFNGEGE 234 | | |

RESULT 1.5

| | |
|--|---|
| ADA10999 | Standard; protein; 259 AA. |
| XX | |
| ADA10999; | |
| AC | |
| XX | |
| DT 06-NOV-2003 | (first entry) |
| XX | |
| DE Human CDNA | differentially expressed in colon cancer #78 product. |
| XX | |
| KW differential expression; colon cancer; cancer; human. | |
| XX | |
| OS Homo sapiens. | |
| XX | |
| PN US2002160382-A1. | |
| XX | |
| PR 11-OCT-2001; 2001US-00981353. | |
| XX | |
| PS 11-OCT-2000; 2000US-0239841P. | |
| XX | |
| PA (LASEK/) LASEK A.W. | |

PA (JONE/) JONES D A.
 XX
 PI Lasek AW, Jones DA;
 XX
 WPI: 2003-265756/26.
 DR N-PSDB; ADA10998.

XX New combination comprising cDNAs that are differentially expressed in
 PT colon disorder, useful for diagnosing, treating, staging or monitoring
 PT treatment for colon cancers.

XX Example 14; SEQ ID NO 117; 231PP; English.

XX The invention relates to a combination comprising cDNAs that are
 CC differentially expressed in colon disorder. The methods and compositions
 CC of the present invention are useful for diagnosing, treating, staging or
 CC monitoring treatment for colon cancer. They are also useful in high
 CC throughput methods for using cDNAs to detect differential expression of
 CC nucleic acids in a sample, screening molecules or compounds to identify a
 CC ligand which specifically binds a cDNA and using a protein to screen
 CC molecules or compounds to identify at least one ligand which specifically
 CC binds the protein. The present sequence represents the amino acid
 CC sequence of a human cDNA differentially expressed in colon cancer
 CC protein.

XX Sequence 260 AA;

| | | | | | | | | |
|----|-----------------------|--|--------------|-------|------------|----|--------|------|
| Qy | Query Match | 61.2% | Score | 788.5 | DB | 6 | Length | 260; |
| Db | Best Local Similarity | 61.6% | Pred. | No. | 1.4e-73; | | | |
| Qy | Matches | 143 | Conservative | 33 | Mismatches | 55 | Indels | 1; |
| Db | Gaps | 1; | | | | | | |
| Qy | 6 | WGYREHNGPIIHWKEFFPIADGDDQSPPIBKTKTKEVKYDQSSLRPLRSIKYDPSAKIISNSGH | 65 | | | | | |
| Db | 5 | WGYKGKANGPERHHWKDFPPIAKGERQSPVDPIDITAKIDPSLKPISVSYDQATSLRILNGH | 64 | | | | | |
| Qy | 66 | SENVDDDTENKSVLRGGGPLTGTSYRLRQVHLHWSADDHGSEHIVDGVSAYAELHVYHWN | 125 | | | | | |
| Db | 65 | AENVEFDSDQDKAVLKGGPLDGTYRLIQFHFTWMSLGQGSSEHTVDCKKYAAELHVHWN | 124 | | | | | |
| Qy | 126 | SDKYPFSFVBAAHEPDGLAVLGVFLQIGEPNSLOQKIDTDLSIKEKGKQTREPTNFDLLS | 185 | | | | | |
| Db | 125 | T-KYGDGFKAQQQPDGLAVLGIPLKVGSAPKSIQKVVDVLDSTKTGKSADETFNFDPRLG | 183 | | | | | |
| Qy | 186 | LPPSWMDYWTPGSLTVPPPLESVTWIVLKOPINISSOOLAKERSLILCTAGE | 237 | | | | | |
| Db | 184 | LPESDLIWTPGSLTVPPPLESVTWIVLKEPISVSSEFQVLFRKLNFGEGE | 235 | | | | | |

Search completed: September 9, 2004, 15:11:24
 Job time : 128 secs


```

! protein - protein search, using sw model
! run on: September 9, 2004, 14:58:44 ; Search time 23 Seconds
! (without alignments)
! 547.869 Million cell updates/sec

title: US-10-069-434-1
refrct score: 229
sequence: 1 M S R L W G Y R E H P N G P I H K E F . . . . . O L A K E R S L C T A E G A A F L 242

```

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

searched: 141681 seconds 52070155 residues

Total number of hits satisfying chosen parameters:

| Post-processing | Minimum DB seq length | Maximum DB seq length | Number of hits |
|--------------------|-----------------------|-----------------------|----------------|
| None | 0 | 200000000 | 100% |
| Exact Match | 0 | 0 | 0% |
| First 45 summaries | 0 | 0 | 0% |

卷之三

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived from analysis of the total score distribution.

| | | | | | | | |
|----|------|-------|-----|---|-------------|---------|-----------------------|
| 1 | 1289 | 100.0 | 262 | 1 | CAHD_HUMAN | Q8n1lq1 | homo sapiens |
| 2 | 1187 | 92.1 | 262 | 1 | CAHD_MOUSE | Q9d6n1 | mus musculus |
| 3 | 848 | 65.8 | 262 | 1 | CAHD_MONDO | Q8hy33 | monodelphis |
| 4 | 808 | 62.7 | 259 | 1 | CAHD_TRIHK | Q8uw55 | tribolodonta |
| 5 | 802 | 62.2 | 260 | 1 | CAHD_HUMAN | P00915 | homo sapiens |
| 6 | 800 | 62.1 | 260 | 1 | CAHD_HORSE | P00917 | equus caballus |
| 7 | 788 | 61.2 | 259 | 1 | CAHD_HUMAN | P00918 | homo sapiens |
| 8 | 788 | 61.2 | 259 | 1 | CAHD_SHEEP | P00922 | ovis aries |
| 9 | 785 | 60.9 | 259 | 1 | CAHD_BOVIN | P00921 | bos taurus |
| 10 | 783 | 60.8 | 260 | 1 | CAHD_BRARE | Q92051 | brachydanius |
| 11 | 782 | 60.7 | 260 | 1 | CAHD_MACMUE | P32517 | macaca nemestrina |
| 12 | 780 | 60.5 | 260 | 1 | CAHD_MACMUE | P00916 | macaca mulatta |
| 13 | 778 | 60.4 | 259 | 1 | CAHD_CHICK | P07630 | gallus gallus |
| 14 | 775 | 60.1 | 260 | 1 | CAHD_SHEEP | P42828 | ovis aries |
| 15 | 774 | 60.0 | 260 | 1 | CAHD_MOUSE | P13634 | mus musculus |
| 16 | 765 | 59.5 | 259 | 1 | CAHD_MOUSE | P16015 | mus musculus |
| 17 | 761 | 59.1 | 259 | 1 | CAHD_RAT | P14141 | rattus norvegicus |
| 18 | 754 | 58.5 | 259 | 1 | CAHD_HORSE | P07450 | equus caballus |
| 19 | 749 | 58.1 | 259 | 1 | CAHD_RABBIT | P00919 | oryctolagus cuniculus |
| 20 | 748 | 58.1 | 259 | 1 | CAHD_RAT | P27139 | rattus norvegicus |
| 21 | 748 | 58.1 | 259 | 1 | CAHD_HUMAN | P07451 | homo sapiens |
| 22 | 747 | 58.0 | 259 | 1 | CAHD_MOUSE | P00920 | mus musculus |
| 23 | 717 | 55.7 | 264 | 1 | CAHD_RABBIT | P07452 | oryctolagus cuniculus |
| 24 | 689 | 53.5 | 264 | 1 | CAHD_HUMAN | P43166 | homo sapiens |
| 25 | 653 | 50.7 | 255 | 1 | CAHD_MOUSE | Q8erq8 | mus musculus |
| 26 | 634 | 49.2 | 317 | 1 | CA5B_MOUSE | Q9gza0 | mus musculus |
| 27 | 631 | 49.0 | 317 | 1 | CA5B_HUMAN | Q9y2d0 | homo sapiens |
| 28 | 589 | 45.7 | 305 | 1 | CAH5_HUMAN | P35218 | homo sapiens |
| 29 | 585 | 45.4 | 299 | 1 | CAH5_MOUSE | P23589 | mus musculus |
| 30 | 581 | 45.1 | 304 | 1 | CAH5_MOUSE | P43165 | mus musculus |
| 31 | 453 | 35.1 | 290 | 1 | CAAH_MOUSE | P28651 | mus musculus |
| 32 | 350 | 35.0 | 289 | 1 | CAAH_HUMAN | P35219 | homo sapiens |
| 33 | 451 | 33.9 | 284 | 1 | CAHC_MOUSE | Q8cje1 | mus musculus |

| | | | | | | | |
|-----|-------|-------|------|-----|---------------|--------|-----------------------|
| 3.4 | 411.5 | 311.9 | 354 | 1 | CABHC_HUMAN | Q43570 | homo sapien |
| | 3.5 | 399 | 355 | 1 | CAHC_RABBIT | Q9Mz02 | oryctolagus cuniculus |
| | 3.6 | 392 | 304 | 1 | CAHH_VACCINIA | P20508 | vaccinia vi |
| | 3.7 | 386 | 29.9 | 304 | CAHH_VACCV | P04195 | vaccinia vi |
| | 3.7 | 386 | 29.9 | 304 | CAHH_VARV | P33065 | variola vir |
| | 3.9 | 376.5 | 29.9 | 337 | CAHE_HUMAN | Q9uLx7 | homo sapien |
| | 4.0 | 369.5 | 29.2 | 337 | CAHE_SHIBEP | P08060 | ovis aries |
| | 4.1 | 367.5 | 28.7 | 307 | CAHE_MOUSE | Q9Wvt6 | mus musculus |
| | 4.2 | 365 | 28.5 | 337 | CAHE_CAMEL | Q10462 | camenorhabdi |
| | 43 | 364.5 | 28.3 | 309 | CAHH_HUMAN | P3280 | homo sapien |
| | 44 | 364.5 | 28.3 | 308 | CAHH_BOVINE | P18915 | bos taurus |
| | 45 | 351 | 27.2 | 319 | CAHH_MOUSE | P18761 | mus musculus |

ALIGNMENTS

RESULT 1

| CAHD | HUMAN | STANDARD: | PRT: |
|------|--|-----------------------------------|---------|
| ID | CAHD_HUMAN | | 262 AA. |
| AC | Q8N1Q1; | (Rel. 42, Created) | |
| DT | 10-OCT-2003 | (Rel. 42, Last sequence update) | |
| DT | 10-OCT-2003 | (Rel. 42, Last annotation update) | |
| DE | Carbonic anhydrase XIII (EC 4.2.1.1) (Carbonic anhydrase XIII (EC 4.2.1.1)) | | |
| DE | <u>Carbonic anhydrase XIII (EC 4.2.1.1) (Carbonic anhydrase XIII (EC 4.2.1.1))</u> | | |

```

>Homo_sapiens (Human)
Homo; sapiens; Craniata; Vertebrata; Euteleostomi;
Eukaryota; Metazoa; Chordata; Mammalia; Buthnesia; Primates; Cattarrhini; Hominidae; Homo.
NCBI_TaxID=9606;
[1] SEQUENCE FROM N.A.
>TTSCTTGCGGAA.

```

Tanigami A., Fujiiwara T., Shibahara T., Goto Y., Hirao M., Shimiz Wakebe H., Hishigaki H., Watanabe T., Ozaki K., Sugiyama Irie R., Otsuka T., Sato H., Wakamatsu A., Ishii S., Yamamoto J., Isono Y., Kawai Hio Y., Saito K., Nishikawa T., Kimura K., Yamashita H., Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Kand Wagatsuma M., Murakawa K., Kanazawa K., Takahashi-Fujii A., Oshim Sugiyama A., Kawakami B., Suzuki Y., Sugano S., Nagahari K.,

"NEDO" human cDNA sequencing project." Submitted (JUL-2002) to the EMBL/GenBank/DDBJ databases.

SEQUENCE FROM N.A.
TISSUE=Pancreas;
MEDLINE=22388257; PubMed=12477932;
Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G., Schatz M.J., Shemmen C.M., Scatliff J.E., Altshuler M., Collins F.S., Wagner L., Shemmen C.M., Schatz M.J., Shemmen C.F., Hsu C.-C., Hopkins R.F., Zeeberg B.R., Butow K.H., Moore T., Max S.I., Wang J.-J., Honig B., Jordan H., Farmer A.A., Rubin G.M., Honig B., Diatchenko L., Marusina K., Soares M.B., Bonaldo M.F., Casarotto T.L., Stapleton M., Brownstein M.J., Usdin T.B., Yoshihara S., Carninci P., Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Bosak S.A., McIwan P.J., McKernan K.J., Malek J.A., Gunzburger M., Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Villalona D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.W., Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G., Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C., Rodriguez A.C., Grimwood J., Schmitz J., Myers R.M., Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalius L., Schein J.E., Jones S.J.M., Marra M.A.; "Generation and initial analysis of more than 15,000 full human and mouse cDNA sequences," Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
- - - FUNCTION: Reversible hydration of carbon dioxide (By catalytic action: $H(2)CO(3) \rightleftharpoons CO(2) + H(2)O$).
- - - COFACTOR: Zinc (By similarity).
- - - SIMILARITY: Belongs to the eukaryotic-type carbonic

| | | | | | | |
|-----------------------|---|--|--------------------|-------------|-----------------------------------|--|
| ID | CAH1 | MONDO | STANDARD; | PRT; | 262 AA. | |
| AC | Q8HYJ3; | | | | | |
| DT | 10-OCT-2003 | (Rel. 42, Created) | | | | |
| DT | 10-OCT-2003 | (Rel. 42, Last sequence update) | | | | |
| DT | 10-OCT-2003 | (Rel. 42, Last annotation update) | | | | |
| DB | Carboxic anhydrase I | (EC 4.2.1.1) (Carbonate dehydratase I) (CA-I). | | | | |
| GN | CAI. | | | | | |
| OS | Monodelphis domestica | (Short-tailed grey opossum). | | | | |
| OC | Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Metatheria; Didelphimorpha; Didelphidae; Monodelphis. | | | | | |
| OC | Mammalia; | | | | | |
| OX | [1] | | | | | |
| RN | | | | | | |
| RP | SEQUENCE FROM N.A. | | | | | |
| RA | Monodelphis domestica (short-tailed grey opossum). | | | | | |
| RT | "Characterisation of opossum (Monodelphis domestica) carbonic anhydrase I and alpha globin coding sequences." | | | | | |
| RT | Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases. | | | | | |
| RL | | | | | | |
| CC | -!- FUNCTION: Reversible hydration of carbon dioxide. | | | | | |
| CC | -!- CATALYTIC ACTIVITY: H(2)CO(3) = CO(2) + H(2)O. | | | | | |
| CC | -!- COFACTOR: Zinc (By similarity). | | | | | |
| CC | -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity). | | | | | |
| CC | -!- SIMILARITY: Belongs to the eukaryotic-type carbonic anhydrase family. | | | | | |
| CC | This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch). | | | | | |
| CC | EMBL: AB056517; BAB3090; 1; -. | | | | | |
| CC | InterPro: IPR001148; Euk_Coanh. | | | | | |
| CC | PFam: PF00194; carb_anhydrase_1. | | | | | |
| CC | ProDom: PDD000865; Euk_CO2_ANHYDRASE_1. | | | | | |
| CC | PROSITE: PS00162; EUK_CO2_ANHYDRASE_1. | | | | | |
| CC | DR | EMBL; InterPro; IPR001148; Euk_Coanh. | | | | |
| CC | DR | PFam: PF00194; carb_anhydrase_1. | | | | |
| CC | DR | ProDom: PDD000865; Euk_CO2_ANHYDRASE_1. | | | | |
| CC | DR | PROSITE: PS00162; EUK_CO2_ANHYDRASE_1. | | | | |
| CC | DR | EMBL; AIA417908; CAD01681; 1; -. | | | | |
| CC | DR | InterPro: IPR001148; Euk_Coanh. | | | | |
| CC | DR | PFam: PF00194; carb_anhydrase_1. | | | | |
| CC | DR | ProDom: PDD000865; Euk_CO2_ANHYDRASE_1. | | | | |
| CC | DR | PROSITE: PS00162; EUK_CO2_ANHYDRASE_1. | | | | |
| KW | Lyase; Zinc. | | | | | |
| FT | INIT | MET | 0 | 0 | BY SIMILARITY. | |
| FT | METAL | 93 | 93 | 93 | ZINC (CATALYTIC) (BY SIMILARITY). | |
| FT | METAL | 95 | 95 | 95 | ZINC (CATALYTIC) (BY SIMILARITY). | |
| FT | METAL | 118 | 118 | 118 | ZINC (CATALYTIC) (BY SIMILARITY). | |
| SQ | SEQUENCE | 259 AA; | 29002 MW; | 29002 MW; | 622013DIC237BBB6 CRC64; | |
| Query Match | 6 | WGYREHNGPITHKEFFPIADGQDQSPTEIKTKVEKYDSSLRPLSKYDPSAKIISNSGH | 65 | | | |
| Best Local Similarity | 62.7% | Score 808.5; | DB 1; | Length 259; | | |
| Matches 150; | Conservative | 64.1%; | Pred. No. 2.1e-63; | | | |
| Matches 150; | Mismatches 32; | Indels 51; | Gaps 1; | | | |
| Db | 4 | WGYADINGPQKWCENPPIANGPQRQSDIDQPTISLDIINNG | 63 | | | |
| Qy | 66 | SPNVDDDDTENKSVLLEGGPLGTSYRQLQVHLWGSADDHGSEHIVGVSYAAELHVHN | 125 | | | |
| Db | 64 | SFQVTADDNDSMTEGPISGKYRKQFIFHWGASDGKSEHTVGCKPAELHVHN | 123 | | | |
| Qy | 126 | SDKYPSEVEAHEPDGLAVLGVEFLOGEPNSQLQKTDLDSIKEKGKQPRFTNFTLDSL | 185 | | | |
| Db | 124 | T-KYASGEANKPDGLAVGVFLQGEDNPKLQKTDLDSIKEKGKQTSFTNFP TCL | 182 | | | |
| Qy | 186 | LPPSWDYWTYPSLITVPPLESVTWVLPKOPINTSSQQLAKFRSLCTAEGEAA | 239 | | | |
| Db | 183 | LPKSLEWXWYPSLITVPPLESVTWVLPKOPINTSSQQLAKFRSLCTAEGEAA | 236 | | | |
| RESULT 5 | | | | | | |
| CAH1 | HUMAN | STANDARD; | PRT; | 260 AA. | | |
| AC | P009415; | | | | | |
| ID | CAH1_HUMAN | | | | | |
| DT | 21-JUL-1986 | (Rel. 01, Created) | | | | |
| DT | 01-APR-1988 | (Rel. 07, Last sequence update) | | | | |
| DT | 15-MAR-2004 | (Rel. 43, Last annotation update) | | | | |
| DB | Carboxic anhydrase I (EC 4.2.1.1) (Carbonate dehydratase I) (CA-I). | | | | | |
| GN | CA1. | | | | | |
| OS | Homo sapiens (Human). | | | | | |
| OC | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. | | | | | |
| CA2. | | | | | | |

RESULT 4

CAH2 TRIKH

STANDARD;

PRT;

259 AA.

AC

Q8UWAS;

ID CAH2_TRIKH

STANDARD;

PRT;

259 AA.

DT

10-OCT-2003

(Rel. 42, Created)

DT

10-OCT-2003

(Rel. 42, Last sequence update)

DT

15-MAR-2004

(Rel. 43, Last annotation update)

DE

Carbonic anhydrase II (EC 4.2.1.1) (Carbonate dehydratase II) (CA-II).

CA2.

- OX NCBI_TaxID=9606;
 RN [1] RP SEQUENCE;
 RX MEDLINE=75091068; PubMed=4217196;
 RA Giraud N., Marriq C., Laurent-Tabusse G.;
 RT "Primary structure of human B erythrocyte carbonic anhydrase. 3.
 RT Sequence of CNBr fragment I and III (residues 149-260).";
 RL Biochimie 56:1031-1043 (1974).
 RN [2] RP SEQUENCE FROM N.A.;
 RX MEDLINE=87174765; PubMed=3104879;
 RA Barlow J.H., Lowe N., Edwards Y.H., Butterworth P.H.W.,
 RT "Human carbonic anhydrase I cDNA";
 RL Nucleic Acids Res. 15:2386-2386 (1987).
 RN [3] RP SEQUENCE FROM N.A.;
 RX MEDLINE=9103039; PubMed=2121614;
 RA Butterworth P.H.W.;
 RA Barlow J.H.M., Barlow J.H., Sowden J.C., Edwards M.,
 RT "Structure and methylation patterns of the gene encoding human
 carbonic anhydrase I. I.;
 RL Gene 93:277-283 (1990).
 RN [4] RP SEQUENCE FROM N.A.
 RC TISSUE=Pancreas, and Spleen;
 RX MEDLINE=22388257; PubMed=12477932;
 RA Krausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shemesh C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J.J., Hsieh P.,
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carrinchi P., Prange C.,
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,
 RA Posak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gundarathne P.H.,
 RA Villalon D.K., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Fahey J., Helton E., Ketteman M., Madan A., Rodriguez S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
 RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smilus D.E.,
 RA Schnier A., Schein J.E., Jones S.J.M., Marrs M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length
 human and mouse cDNA sequences";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
 RN [5] RP SEQUENCE OF 19-260;
 RX MEDLINE=12241008; PubMed=4625868;
 RA Andersson B., Nyman P.O., Sirrid L.;
 RT "Antino acid sequence of human erythrocyte carbonic anhydrase B. 2.";
 RL Biochem. Biophys. Res. Commun. 48:670-677 (1972).
 RN [6] RP SEQUENCE OF 11-260;
 RX MEDLINE=3134579; PubMed=4632246;
 RA Lin K.-T.D., Deutrich H.F.;
 RT "Human carbonic anhydrases. XI. The complete primary structure of
 carbonic anhydrase B.";
 RL J. Biol. Chem. 248:1885-1893 (1973).
 RN [7] RP REVISIONS;
 RX MEDLINE=4143468; PubMed=4207120;
 RA Lin K.-T.D., Deutrich H.F.;
 RT "Human carbonic anhydrases. XII. The complete primary structure of
 the C isozyme.";
 RL J. Biol. Chem. 249:2329-2337 (1974).
 RN [8] RP X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS);
 RX MEDLINE=15120492; PubMed=804711;
 RA Kaaman K.K., Notstrand B., Fridborg K., Loevgren S., Ohlsson A.,
 RA Petef M.;
 RT "Crystal structure of human erythrocyte carbonic anhydrase B. Three-
 dimensional structure at a nominal 2.2-A resolution.";
 RT
- RL RN [9] Proc. Natl. Acad. Sci. U.S.A. 72:51-55 (1975).
 RN RP VARIANT GUAM;
 RX MEDLINE=81130805; PubMed=6781336;
 RA Omoto K., Ueda S., Goriki K., Takahashi N., Misawa S., Pagaran I.G.;
 RT "Population genetic studies of the Philippine Negritos. III.
 RT Identification of the carbonic anhydrase-I variant with CA1 Guam.";
 RL Am. J. Hum. Genet. 33:105-111 (1981).
 RN RN [10] VARIANT MICHIGAN-1.
 RP MEDLINE=95170740; PubMed=78666410;
 RX MEDLINE=95170740; PubMed=78666410;
 RA Chegwidden W.R., Wagner L.E., Venta P.J., Bergenhem N.C.H.,
 RA Yu Y.-S.L., Tashian R.E.;
 RT "Marked zinc activation of ester hydrolysis by a mutation, 67-His
 (CAT) to Arg (CCT), in the active site of human carbonic anhydrase
 I.";
 RT Hum. Mutat. 4:294-296 (1994).
 RL RN [11] FUNCTION: Reversible hydration of carbon dioxide.
 CC -!- CATALYTIC ACTIVITY: H(2)CO(3) = CO(2) + H(2)O.
 CC -!- COFACTOR: Zinc.
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
 CC -!- SIMILARITY: Belongs to the eukaryotic-type carbonic anhydrase
 family.
 CC
- CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement. (See <http://www.ebi-sib.ch/announce/>
 CC or send an email to license@ebi-sib.ch).
 CC
- CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement. (See <http://www.ebi-sib.ch/announce/>
 CC or send an email to license@ebi-sib.ch).
 CC
- DR X05014; CAA28663; 1 - .
 DR EMBL; M33987; AAA51910; 1 - .
 DR EMBL; BC027590; AAH27890; 1 - .
 DR PIR; JQ0786; CRHUL.
 DR PDB; 2CAB; 17-JUL-94.
 DR PDB; 1AZM; 30-APR-94.
 DR PDB; 1BZM; 30-APR-94.
 DR PDB; 1CZM; 30-APR-94.
 DR PDB; 1HC8; 30-APR-94.
 DR PDB; 1HUG; 30-APR-94.
 DR PDB; 1IHH; 30-APR-94.
 DR PDB; 1CRM; 07-FEB-95.
 DR PDB; 1J9W; 13-JUN-01.
 DR PDB; 1JU0; 17-JUN-03.
 DR PMDB; 2DPAGE; P00915; - .
 DR Genew; HGNC; 1336B; CAL.
 DR MTM; 114800; .
 DR GO:0004089; F:carbonate dehydratase activity; TAS.
 DR InterPro; IPR001148; Euk Coanhd.
 DR PFAM; PF00194; carb_anhydrase_1.
 DR PRODOM; PD000865; Euk_Coanhd; 1.
 DR PROSITE; PS00162; EUK_CO2_ANHYDRASE_1.
 KW Lyase; Zinc; Metal-binding; Acetylation; Polymorphism; 3D-structure.
 FT INT-MET 0 0 ACETYLYATION.
 FT MOD-RES 1 1 ZINC (CATALYTIC).
 FT METAL 94 94 ZINC (CATALYTIC).
 FT METAL 96 96 ZINC (CATALYTIC).
 FT METAL 119 119 ZINC (CATALYTIC).
 FT VARIANT 67 67 H -> R (in Michigan-1).
 DR /FTD=VAR 001378.
 DR G -> R (in Guam).
 DR DN -> ND (IN REF. 1, 4 AND 5).

| | |
|-----------------------|--|
| RT | "Sequence of the low activity equine erythrocyte carbonic anhydrase and delineation of the amino acid substitutions in various polymorphic forms"; J. Biol. Chem. 255:19196-9204 (1980). |
| RL | -!- CATALYTIC ACTIVITY: Reversible hydration of carbon dioxide. |
| CC | -!- COFACTOR: Zinc. |
| CC | -!- SUBCELLULAR LOCATION: Cytoplasmic. |
| CC | -!- POLYMORPHISM: The sequence shown is that of the electrophoretically homogeneous D isozyme. |
| CC | -!- SIMILARITY: Belongs to the eukaryotic-type carbonic anhydrase family. |
| CC | PIR; A01140; CRH01D. |
| DR | HSSP; P00915; IHC8. |
| DR | InterPro; IPR001148; Euk_CoAhd. |
| PFam | PF00194; carb_anhydrase; 1. |
| DR | ProDom; PD000865; Euk_CoAhd; 1. |
| DR | PROSITE; PS00162; Euk_CO2_Anhydrase; 1. |
| DR | PROSITE; PS00162; Euk_CoAhd; 1. |
| KW | Lyase; Zinc; Metal-binding; Polymorphism. |
| FT | ZINC (CATALYTIC). |
| FT | METAL; Zinc (CATALYTIC). |
| FT | METAL; Zinc (CATALYTIC). |
| FT | METAL; Zinc (CATALYTIC). |
| FT | VARIANT; A -> D (IN THE T ISOZYME). |
| FT | VARIANT; S -> G (IN HOMOGENEOUS D ISOZYME). |
| FT | VARIANT; DGP -> GCF (IN A2 ISOZYME). |
| FT | VARIANT; S -> H (IN HOMOGENEOUS D ISOZYME). |
| FT | VARIANT; L -> G (IN HOMOGENEOUS D ISOZYME). |
| FT | VARIANT; R (IN THE A1 AND B ISOZYMES). |
| FT | VARIANT; S -> R (IN HOMOGENEOUS D ISOZYME). |
| FT | VARIANT; C -> Y (IN HOMOGENEOUS D ISOZYME). |
| FT | VARIANT; Q -> R (IN THE B ISOZYME). |
| FT | VARIANT; Q -> A (IN HOMOGENEOUS D ISOZYME). |
| SQ | SEQUENCE 260 AA; 28895 MW; 04384FC5F1FAA48 CRC64; |
| Query Match | 62.1%; Score 800; DB 1; Length 260; |
| Best Local Similarity | 62.0%; Pred. No. 1.2e-52; |
| Matches | 35; Mismatches 54; Indels 0; Gaps 0; |
| CC | Best Local Similarity 60.3%; Pred. No. 7.9e-63; |
| CC | Conservative 43; Mismatches 50; Indels 0; Gaps 0; |
| CC | Score 802; DB 1; Length 260; |
| CC | Matches 141; Conservative 43; Mismatches 50; Indels 0; Gaps 0; |
| Qy | 6 WGYRENGPIKKEFPIPLADGDOQSPSLEIKTKREVKYDSLRPLSIKYDPSSAKIISNSGH 65 |
| Db | 5 WGYDNGPQEWKSYLPIANGNQSPDVIKTSKHTSLKPISVNPATAEITVNH 64 |
| Qy | 6 SFNVDPPDTENKSVLRGPLTGSYRLRQVHLWGSADDHGSEHTIVDGYSAAELHHVWN 125 |
| Db | 65 SFHVNPDNDNEVSVLKGQPFSDYRLFPHFMGSTNEHGSEHTDVGKYSAAELHHVWN 124 |
| Qy | 126 SDKYPSTVEAAHPDGIAVIGYFLQIGEPNSQLOKTTDLSIKEKGKQTRTNFDLISL 185 |
| Db | 125 SKYSSFDESDQADGLIILGKVGEANPKLQVLDAAEVKTKGGKKAPPKNFDFSSL 184 |
| Qy | 186 LPPSWDYWTYFGSLTVPPJLIESVTWLVKOPINISSSQLAKERSLCTAEGEEA 239 |
| Db | 185 LPSSPDWYTYSGLTHPPJLIESVTWLVCKENISISSQLQFRSLLSNVSEGKA 238 |

| RESULT 6 | | | | | |
|------------------|--|---------------------------------|------|---------|---|
| CAHAN | CAH1_HORSE | STANDARD; | PRT; | 260 AA. | |
| ID | CAH1 | HORSE | | | |
| AC | P00917; | | | | |
| DT | 21-JUL-1986 | (Rel. 01, Created) | | | |
| | | | | | Last sequence update |
| DT | 21-JUL-1986 | (Rel. 01, Last sequence update) | | | |
| | | | | | 10-OCT-2003 (Rel. 42, Last annotation update) |
| DE | Carbonic anhydrase I (EC 4.2.1.1) (Carbonate dehydratase I) | | | | |
| | | | | | Carbonic anhydrase C. |
| CAI. | | | | | |
| GN | Equus caballus (Horse). | | | | |
| | Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Perissodactyla; Equidae; Equus. | | | | |
| OS | | | | | |
| OC | | | | | |
| NCBI_TAXID=9796; | | | | | |
| RN | [1] | | | | |
| RP | SEQUENCE (ISOZYME D). | | | | |
| RX | MEDLINE=8106999; PubMed=6773961; | | | | |
| RA | Jabusch J.R.; Bray R.P.; Deutsch H.F.; Lin K.-T.D.; Deutscher P.; Deutsch H.F.; | | | | |
| RT | "Human carbonic anhydrates". XII. The complete primary structure of carbonic anhydrase II (EC 4.2.1.1) (Carbonate dehydratase II) (CA-II). | | | | |
| CAH2 | HUMAN | STANDARD; | PRT; | 259 AA. | |
| ID | CAH2 | HUMAN | | | |
| AC | P00918; Q96ET9; | | | | |
| DT | 21-JUL-1986 (Rel. 01, Created) | | | | |
| | | | | | Last sequence update |
| DT | 21-JUL-1986 (Rel. 01, Last annotation update) | | | | |
| DE | Carbonic anhydrase II (EC 4.2.1.1) (Carbonate dehydratase II) (CA-II). | | | | |
| CA2. | | | | | |
| GN | Homo sapiens (Human). | | | | |
| OS | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Buteleostomi; Mammalia; Eutheria; Primates; Hominidae; Homo. | | | | |
| OC | | | | | |
| NCBI_TAXID=9606; | | | | | |
| RN | [1] | | | | |
| RP | SEQUENCE. | | | | |
| RX | MEDLINE=74143468; PubMed=4207120; | | | | |
| RA | Deutsch H.F.; Lin K.-T.D.; Deutscher P.; Deutsch H.F.; | | | | |
| RT | "Human carbonic anhydrates". XII. The complete primary structure of carbonic anhydrase II (EC 4.2.1.1) (Carbonate dehydratase II) (CA-II). | | | | |

- RT the C isozyme." ;
 RL J. Biol. Chem. 249:2329-2337(1974).
 RN
 RP [2]
 SEQUENCE.
 RX MEDLINE=77006079; PubMed=823150;
 RA "Primary structure of human carbonic anhydrase C." ;
 RT
 RL J. Biol. Chem. 251:5457-5463(1976).
 RN
 RP SEQUENCE FROM N.A. ;
 RX MEDLINE=87231043; PubMed=3108857;
 RA "Nucleotide sequence of human liver carbonic anhydrase III cDNA." ;
 RT
 RL
 RN
 RP SEQUENCE FROM N.A. ;
 RX MEDLINE=88005190; PubMed=3121496;
 RA "Cloning, expression, and sequence homologies of cDNA for human
 carbonic anhydrase II." ;
 RT
 RL Genomics 1:159-166(1988).
 RN
 RP TISSUE-Ovary;
 RX MEDLINE=22388257; PubMed=12477932;
 RA Strausberg R.L., Feingold F.A., Grouse L.H., Derge J.G., Shemerlin C.M., Schubler G.D., Klausner R.D., Collins P.S., Wagner L., Sherman C.M., Schaefer C.F., Bhat N.K., Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K., Hsieh F., Hollings R.F., Jordan H., Moore T., Wang J., Hsieh F., Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L., Stapleton M., Soares M.B., Bonaldo M.P., Casavant T.L., Scheetz T.E., Brownstein M.J., Udsin T.B., Yoshiaki S., Carrinchi P., Prange C., Brown S.S., Logueillo N.A., Peters G.J., Abramson R.D., Mullally S.J., Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H., Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W., Villalon D.K., Munny D.M., Sodergren E.J., Lu X., Gibbs R.A., Fahey J., Helton E., Ketteman M., Madan A., Rodriguez S., Sanchez A., Whiting M., Madan A., Young A.C., Shevchenko Y., Boutlard G.G., Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C., Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E., Schnarch A., Schein J.E., Jones S.J.M., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences";
 RT Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
 RN
 RP SEQUENCE OF 1-76 FROM N.A. ;
 RX MEDLINE=86077780; PubMed=3000449;
 RA Venter P.J., Montgomery J.C., Hewett-Emmett D., Tashian R.E.;
 RT "Comparison of the 5' regions of human and mouse carbonic anhydrase II genes and identification of possible regulatory elements." ;
 RN
 RP X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS). ;
 RX MEDLINE=72111787; PubMed=4621826;
 RA Strandbergh B., Carlbom U., Jaerup L., Loeyven S., Petef M.;
 RT "Crystal structure of human carbonic anhydrase C." ;
 RN
 RP X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS). ;
 RX MEDLINE=89315726; PubMed=3151019;
 RA Eriksson A.E., Jones T.A., Liljas A.;
 RT "Refined structure of human carbonic anhydrase II at 2.0 Å resolution." ;
 RL Proteins 4:274-282 (1988).
 RN
 RP X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS). ;
 RX MEDLINE=59315727; PubMed=3151020;
 RA Eriksson A.E., Kyllonen P.M., Jones T.A., Liljas A.;
 RT "Crystallographic studies of inhibitor binding sites in human carbonic anhydrase II: a pentacoordinated binding of the SCN-ion to the zinc at high pH." ;

- RL Proteins 4:283-293 (1988).
 RN RN [10]
 RP X-RAY CRYSTALLOGRAPHY (2.25 ANGSTROMS).
 RX MEDLINE=59315728; PubMed=554186;
 RA Stams T., Chen Y., Borack-Sjodin P.A., Hurt J.D., Liao J., May J.A., Den T., Laips P., Silverman D.N., Christianson D.W.;
 RT "Structures of murine carbonic anhydrase IV and human carbonic anhydrase II complexed with brinzolamide: molecular basis of isoform discrimination." ;
 RL Protein Sci. 7:556-563 (1998).
 RN RN [11]
 RP X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS).
 RX MEDLINE=20510299; PubMed=11076507;
 RA Cox J.D., Hunt J.A., Compier K.M., Pierke C.A., Christianson D.W.;
 RT "Structural influence of hydrophobic core residues on metal binding and specificity in carbonic anhydrase II." ;
 RL Biochemistry 39:13687-13694 (2000).
 RN RN [12]
 RP VARIANT JOGJAKARTA. ;
 RX MEDLINE=83100296; PubMed=6817747;
 RA Jones G.L., Sofro A.S.M., Shaw D.C.;
 RT "Chemical and enzymological characterization of an Indonesian variant of human erythrocyte carbonic anhydrase II, CAII. Jogyjakarta (17 Lys leads to Glu)." ;
 RL Biochem. Genet. 20:979-1000 (1982).
 RN RN [13]
 RP VARIANT MELBOURNE. ;
 RX MEDLINE=83226368; PubMed=6407977;
 RA Jones G.L., Shaw D.C.;
 RT "A chemical and enzymological comparison of the common major human erythrocyte carbonic anhydrase II, its minor component, and a new genetic variant, CA II Melbourne (237 Pro leads to His)." ;
 RL Hum. Genet. 63:392-399 (1983).
 RN RN [14]
 RP VARIANT CA2 DEFICIENCY TYR-106. ;
 RX MEDLINE=92026087; PubMed=1928091;
 RA Venta P.J., Welty R.J., Johnson T.M., Sly W.S., Tashian R.E.;
 RT "Carbonic anhydrase II deficiency syndrome in a Belgian family is caused by a point mutation at an invariant histidine residue (107 His-->Tyr): complete structure of the normal human CA II gene." ;
 RL Am. J. Hum. Genet. 49:1082-1090 (1991).
 RN RN [15]
 RP VARIANT CA2 DEFICIENCY TYR-106. ;
 RX MEDLINE=9-2119270; PubMed=1542674;
 RA Roth D.E., Venta P.J., Tashian R.E., Sly W.S.;
 RT "Molecular basis of human carbonic anhydrase II deficiency." ;
 RL Proc. Natl. Acad. Sci. U.S.A. 89:1804-1808 (1992).
 RN RN [16]
 RP VARIANT CA2 DEFICIENCY TYR-106. ;
 RX MEDLINE=9-6431156; PubMed=8934238;
 RA Soga H., Yukizane S., Yoshida I., Koga Y., Aramaki S., Kato H.;
 RT "A point mutation in exon 3 (His 107-->Tyr) in two unrelated Japanese patients with carbonic anhydrase II deficiency system involvement." ;
 RL Hum. Mutat. 9:383-387 (1997).
 RN RN [17]
 RP VARIANT CA2 DEFICIENCY PRO-91. ;
 RX MEDLINE=9-72881992; PubMed=9143911;
 RA Hu P.Y., Lim E.J., Ciccone J.J., Strisciuglio P., Sly W.S.;
 RT "Seven novel mutations in carbonic anhydrase II deficiency syndrome identified by SSCP and direct sequencing analysis." ;
 RL Hum. Mutat. 9:383-387 (1997).
 CC -1- FUNCTION: Reversible hydration of carbon dioxide.
 CC -1- CATALYTIC ACTIVITY: $H_2CO(3) = CO(2) + H_2O$.
 CC -1- COFACTOR: Zinc.
 CC -1- SUBCELLULAR LOCATION: Cyttoplasmic.
 CC -1- DISEASE: Defects in CA2 are the cause of carbonic anhydrase II deficiency (CA2 deficiency) [MIM#259730]. It can be associated with osteopetrosis, renal tubular acidosis, cerebral calcification (mable brain disease) and mental retardation. Inheritance is autosomal dominant.
 CC -1- SIMILARITY: Belongs to the eukaryotic-type carbonic anhydrase family.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the ENB outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

[1] SEQUENCE.
RN MEDLINE=75054988; PubMed=4215456;
RX TANIS R.J.; FERRELL R.E.; TASHIAN R.E.;
RA "Amio acid sequence of sheep carbonic anhydrase C.";
RT Biophys. Acta 371:534-548 (1974).
RL [2]
RN [2]
RP VARIANT FORM.
RX MEDLINE=79145512; PubMed=106835;
RA Mallett B.; Guiian J.M.; Sciaky M.; Charrel M.;
RT "Multiple molecular forms of erythrocyte carbonic anhydrase of
sheep";
RL Biochim. Biophys. Acta 576:290-304 (1979).
CC -|- CATALYTIC ACTIVITY: H(2)CO(3) = CO(2) + H(2)O.
CC -|- COFACTOR: Zinc.
CC -|- SUBCELLULAR LOCATION: Cytoplasmic.
CC -|- MISCELLANEOUS: ONE MINOR AND THREE MAJOR FORMS WERE ISOLATED
CC -|- CHROMATOGRAPHICALLY.
CC -|- SIMILARITY: Belongs to the eukaryotic-type carbonic anhydrase
CC Family.
DR PIR; A01145; CRSH2.
DR HSSP; P00918; 1C1M.
DR InterPro; IPR01148; Euk_Coanh.
DR Pfam; PF00194; carb_anhydrase_1.
DR Prodrom; P0000865; Euk_Coanh_1.
DR PROSITE; PS00122; EUR_CCO_ANHYDRASE_1.
DR Lyase; Zinc; Acetylation.
KW MOD RES 1 1
FT METAL 93 93
FT METAL 95 95
FT METAL 118 118
FT VARIANT 35 X (IN ONE OF THE MAJOR FORMS).
SQ SEQUENCE 259 AA; 29080 MW; E0BDD6B67263604 CRC64;
Query Match 61.2%; Score 788.5; DB 1; Length 259;
Best Local Similarity 62.1%; Pred. No. 1.2e-61;
Matches 144; Conservative 39; Mismatches 48; Indels 1; Gaps 1;

Query Match 61.2%; Score 788.5; DB 1; Length 259;
Best Local Similarity 61.6%; Pred. No. 1.2e-61;
Matches 143; Conservative 33; Mismatches 55; Indels 1; Gaps 1;
Qy 6 WGYREANGPIHWKEFFIADGQQSPTEIKTKVEKYDSSRPLSIKDPSSAKIISNSGH 65
Db 4 WGYGRNGPPIHWKDPINKERQSPVTDIPTAKDPSLKEPKRPSVYDQNSLRLINGH 63
Qy 66 SENVDDDTENKSVLRGGLPTGSYRLRQVILHWSADDHGSEHIEVGVSTAAELHVHN 125
Db 64 AFNVEFDSDQDAVLRKGPLGTYRLQFHFGSDLGQGSHTDVKKKAELHVHN 123
Qy 126 SDKYPSPVEAHEPDGLAVLGYLOQEPNSLQLQKTDTSIKEKGKQTFTNFEDLIS 185
Db 124 T-KYGFDTKAQQPDGLAVLGYFLKVGSAKPLQKVTDLSIKTGKSAFTNFEDRGL 182
Qy 186 LPPSWDYWTVPEGSLLTPPLLESVTWVILKOPINISSQLAKEFSLCLTAGEE 237
Db 183 LKRALNYWTVPGSLTNPPLLESVTWVILKEPTSVSSQMLKFRSLNFNAESE 234
Qy 183 LPESLWYWTVPGSLTVPPPLLESVTWVILKEPTSVSSQMLKFRSLNFNAESE 237
Db 183 LPESLWYWTVPGSLTVPPPLLESVTWVILKEPTSVSSQMLKFRSLNFNAESE 234

RESULT 9
CAH2_BOVIN STANDARD PRT; 259 AA.
ID CAH2_BOVIN STANDARD PRT; 259 AA.
AC P00925; (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DB Carbonic anhydrase II (EC 4.2.1.1) (Carbonate dehydrate II) (CA-II).
CA2. OS Bos taurus (Bovine).
OC Mammalia; Eutheria; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Bovidae; Bovine; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC Mammalia; Eutheria; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Bovidae; Bovine; Bos.
NCBI_TaxID=9913;
RN [1]
RP SEQUENCE.
RC TISSUE=Erythrocyte;

RX MEDLINE=77065798; PubMed=826282;
 RA Sciairy M.; Limozin N.; Filippi Foveau D.; Gulian J.M.,
 RA Laurent-rabasse G.;
 RT "Primary structure of bovine erythrocyte carbonic anhydrase homologue
 CI. II. Complete sequence.";
 RT Biochimie 58:1071-1082(1976).
 RL [2]
 RN REVISIONS.
 RP MEDLINE=77242599; PubMed=19093;
 RA Gulian J.M.; Linozin N.; Mallet B.; di Costanzo J.; Charrel M.;
 RT "Genetic independence of two forms of carbonic anhydrase from bovine
 erythrocytes.";
 RL Biochimie 59:293-302(1977).
 CC -!- FUNCTION: Reversible hydration of carbon dioxide.
 CC -!- CATALYTIC ACTIVITY: H(2)CO(3) = CO(2) + H(2)O.
 CC -!- COFACTOR: Zinc (By similarity).
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
 CC -!- MISCELLANEOUS: ONE MINOR AND TWO MAJOR FORMS WERE ISOLATED
 CC CHROMATOGRAPHICALLY.
 CC -!- SIMILARITY: Belongs to the eukaryotic-type carbonic anhydrase
 family.
 CC -!- DATABASE: NAME=Worthington enzyme manual;
 CC WWW="http://www.worthington-biochem.com/CA/".
 DR PIR; AC1144; CRB02.
 DR PDB; 1G6V; 22-NOV-00.
 DR InterPro; IPR001148; Euk Coanh.
 DR Pfam; PF00194; Carb anhydrase; 1.
 DR ProdDom; PD000865; Euk Coanh.
 DR PROSITE; PS00162; EUK CO2 ANHYDRASE; 1.
 KW Lyase; Zinc; Acetylation; 3D-structure.
 FT MOD_RES 1 1
 FT METAL 93 93
 FT METAL 95 95
 FT METAL 118 118
 FT VARIANT 56 56 R -> Q (IN ONE OF THE MAJOR FORMS).
 SQ SEQUENCE 259 AA; 28980 MW;
 Query Match Score 785.5 ; DB 1 ; Length 259;
 Best Local Similarity 61.2%; Pred. No. 2.2e-61 ;
 Matches 142 ; Conservative 41 ; Mismatches 48 ; Indels 1 ;
 Gaps 1 ;
 DB 4 WGYGKHBQZWHWRKDFTIANGEROVSPIVNDTAKWDALKPLALVGEATSRMVNNGH 63
 DB 66 SENVYDDDTENKSYLRGGPLTGSYRLQVHLLGSSADHGSEHITDVGSYAAELHVHN 125
 DB 64 SENVYDDDSQDKAVLKDGPLTGTYRLQVQFHEHGGSSBHQGSEHTVDKRYAAELHVHN 123
 Qy 126 SDKYPSPVEAAHEDGLAVLGVFLQIGEBSNQLOKITHLDLSKEKGKQTTRTFNFDLISL 1.85
 DB 124 T-KYGFDTAAQPDGLAVGVFLKVGANPAQKVLDAISITKGKSTDFFNFDFGSLL 182
 Qy 186 LPPSWDWTYPGSLTVPPPLESIVLWIKOPINTSSQQLAKFRSLCTAEGE 237
 DB 183 LPNVLDWTYPGSLTVPPPLESIVLWIKEPISVSSQMLKFRTLNFAEGE 234
 Qy 186 LPPSWDWTYPGSLTVPPPLESIVLWIKOPINTSSQQLAKFRSLCTAEGE 237
 DB 183 LPNVLDWTYPGSLTVPPPLESIVLWIKEPISVSSQMLKFRTLNFAEGE 234
 RESULT 11
 CAH1 MACNE STANDARD ; ERT ; 260 AA .
 ID CAH1 MACNE
 AC P35217;
 DT 01-FEB-1994 (Rel. 28, Created)
 DT 01-FEB-1994 (Rel. 28, Last sequence update)
 AC Q2051;
 DT 01-NOV-1997 (Rel. 35, Created)
 DR 15-JUL-1998 (Rel. 36, Last sequence update)
 DR Carbonic anhydrase (EC 4.2.1.1) (Carbonate dehydratase 1) (CA-I).
 GN CA1.
 OS Macaca nemestrina (Pig-tailed macaque).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Buteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecidiae;
 OC Cercopithecinae; Macaca.
 NCBI_TaxID=545;
 RN SEQUENCE FROM N.A.
 RP MEDLINE=95137401; PubMed=7835713;
 RX NCBI_TaxID=7955;
 RA Hopkins P.J.; Bergenheim N.C.H.; Venta P.J.; Hewett-Emmett D.,

SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.

MEDLINE=9750269; PubMed=1089083;

RA Peterson R.E.; Tu C.; Linne P.J.;

RT "Isolation and characterization of a carbonic anhydrase homologue
 from the zebrafish (*Danio rerio*).";

RL J. Mol. Evol. 44:432-439(1997).

CC -!- FUNCTION: Reversible hydration of carbon dioxide.

CC -!- CATALYTIC ACTIVITY: H(2)CO(3) = CO(2) + H(2)O.

CC -!- COFACTOR: Zinc (By similarity).

CC -!- SIMILARITY: Belongs to the eukaryotic-type carbonic anhydrase
 family.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial/

CC entities requires a license agreement. (See http://www.isb-sib.ch/announce/

CC or send an email to license@isb-sib.ch).

CC EMBL; U55177; AAB82303; 1; -.

DR PIR; 108463; T08463.

DR HSSP; P00918; 1CIM.

DR ZDB-CENE-080526-39; cahz.

DR InterPro; IPR00114; Euk Coanh.

DR Pfam; PF00194; carb_Coanhase; 1.

DR ProdDom; PD000865; Euk_Co2_Anhydase; 1.

DR PROSITE; PS00162; Euk_CO2_ANHYDRASE; 1.

KW Lyase; Zinc; Acetylation; 3D-structure.

FT MOD_RES 1 1
 ACETYLATION.FT METAL 93 93
 ZINC (CATALYTIC).FT METAL 95 95
 ZINC (CATALYTIC).FT METAL 118 118
 ZINC (CATALYTIC).

FT VARIANT 56 56 R -> Q (IN ONE OF THE MAJOR FORMS).

SQ SEQUENCE 259 AA; 28980 MW;

Query Match Score 783.5 ; DB 1 ; Length 260;

Best Local Similarity 61.8%; Pred. No. 3.3e-61 ;

Matches 144 ; Conservative 34; Mismatches 54; Indels 1 ; Gaps 1 ;

Qy 5 SWGYREHNGPIHKWEFFPIADGQDQSPTIEKTKEYVYDQSSLRPLSIKYDQSSAKIIISNGH 64

Db 4 AWGYGPAGDPGEWSAEPIANGPRQSPIDIVTAQQAIDPSLKHKLKDYDPAIKSILNNG 63

Qy 65 HSPNVDDEDTENKSVLRGGLPTGSYRLQVHLLGHSSADDHGSEHITDVGSYAAELHVHW 124

Db 64 HSPQVDFDDDSNTSLAGGPITGTYRLQFHPHGSSDPKGSEHTIACTKPOELHLVHW 123

Qy 65 NSDKYPSVVEAAHEDGLAVLGVFLQIGEBSNQLOKITHLDLSKEKGKQTTRTFNFDLISL 1.84

Db 64 NT-KYPNFGEAASKPDGIAVGVFLKGAANPRLQVKYLDADDIKSGROTTFANFPEK 182

Qy 185 LPPSWDWTYPGSLTVPPLESIVLWIKOPINTSSQQLAKFRSLCTAEGE 237

Db 183 LIPASLDWTYEGSLTTPPLESIVLWIKEPISVSSQMLKFRTLNFAEGE 235

RESULT 11

CAH1 MACNE STANDARD ; ERT ; 260 AA .

ID CAH1 MACNE

AC P35217;

DT 01-FEB-1994 (Rel. 28, Created)

DT 01-FEB-1994 (Rel. 28, Last sequence update)

AC Q2051;

DT 10-OCT-2003 (Rel. 42, Last annotation update)

DE Carbonic anhydrase I (EC 4.2.1.1) (Carbonate dehydratase 1) (CA-I).

GN CA1.

OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Buteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecidiae;

OC Cercopithecinae; Macaca.

NCBI_TaxID=545;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=95137401; PubMed=7835713;

RA Hopkins P.J.; Bergenheim N.C.H.; Venta P.J.; Hewett-Emmett D.,

RN

Tashian R.E.;
 RA "Characterization of the gene encoding carbonic anhydrase I from the
 RT pigtail macaque";
 RT Gene 15:265-269 (1995)
 RL -!- FUNCTION: Reversible hydration of carbon dioxide.
 CC -!- CATALYTIC ACTIVITY: $\text{H}_2\text{CO}(3) = \text{CO}(2) + \text{H}_2\text{O}$.
 CC -!- COFACTOR: Zinc.
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
 CC -!- POLYMORPHISM: Four electrophoretic alleles are known to exist, they
 CC are designated A (shown here), B, C and D.
 CC -!- SIMILARITY: Belongs to the eukaryotic-type carbonic anhydrase
 CC family.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL / L25082; AAA65198.1; - .
 HSSP: P00915; IJZM.
 InterPro: IPR01148; Euk_Coanhds.
 Pfam: PF00194; carb_anhydride_1.
 Prodrom: P000085; Euk_Coanhds_1.
 PROSITE: PS000152; Euk_CO2_ANHYDRASE_1.
 Lyase / Zinc: Metal-binding; Acetylation; Polymorphism.
 INIT / MET 0 0 BY SIMILARITY.
 MOD / EFS 1 1 ACETYLATION (BY SIMILARITY)
 MOD / MET 0 0

| | | | | | | | | |
|----------|---------|-----------|------------------|---------------------|--|--|--|--|
| METAL | 94 | 94 | ZINC | (CATALYTIC) | | | | |
| METAL | 96 | 96 | ZINC | (CATALYTIC) | | | | |
| METAL | 119 | 119 | ZINC | (CATALYTIC) | | | | |
| VARIANT | 241 | 242 | SQ | -> IE (IN ALLELE B) | | | | |
| SEQUENCE | 260 AA: | 28807 MW: | 04FDFFEA9F0C85E0 | CRC64; | | | | |

Match Score 782; DB-61; Length 260;
 Local Similarity 58.6%; Pred. 4.4e-61;
 Conservative 47; Mismatches 49; Indels 0; Gaps 0;

Local Similarity 58.6%; Pred. 4.4e-61;
 Conservative 47; Mismatches 49; Indels 0; Gaps 0;

Local Similarity 58.6%; Pred. 4.4e-61;
 Conservative 47; Mismatches 49; Indels 0; Gaps 0;

Local Similarity 58.6%; Pred. 4.4e-61;
 Conservative 47; Mismatches 49; Indels 0; Gaps 0;

Local Similarity 58.6%; Pred. 4.4e-61;
 Conservative 47; Mismatches 49; Indels 0; Gaps 0;

123 SAKIOSSHEA SKHODGAVIY ULMKU SEANPAQKU QALHALIKAGARAF INF DF SIL 184
186 LPPSWD YWWTYP DGSILTY PPLLESVTWIVLKQPINISSQQLAKRSLLCTAGE 237
195 1DCEST DEWWTYCCSITMDT YCFCVTTWV CEEPSYCSMUSLQDTCI LQDTCI LQDTCI 225

RESULT 12
 CAH1_MACMU MACMU STANDARD; PRT; 260 AA.
 AC P00916;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 21-JUL-1986 (Rel. 01, Last annotation update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Carbonic anhydrase I (EC 4.2.1.1) (Carbonate dehydratase I) (CA-I).
 GN CAI.
 OS Macaca mulatta (*Rhesus macaque*).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Butheria; Primates; Catarrhini; Cercopithecoidea;
 OC Cercopithecinae; Macaca.
 NCBI_TaxID=9544;
 NCBI_TaxID=9031;
 RN [1] _
 RP SEQUENCE FROM N.A.
 RC STRAIN=Hubbard White Mountain; TISSUE=Testis;
 MEDLINE#=95011620; PubMed=7226806;
 RX
 RA Mezquita J., Pau M., Mezquita C.;
 RT "A novel carbonic anhydrase II mRNA isolated from mature chicken
 testis displays a TATA box and other promoter sequences in a leader
 5' untranslated region not present in somatic tissues.";
 RT Gene 147:231-235 (1994).
 RN [2] _
 RP SEQUENCE FROM N.A.
 RC STRAIN=White leghorn;
 MEDLINE#=87146391; PubMed=3029691;
 RX
 RA Mezquita J., Pau M., Mezquita C.;
 RT "A novel carbonic anhydrase II mRNA isolated from mature chicken
 testis displays a TATA box and other promoter sequences in a leader
 5' untranslated region not present in somatic tissues.";
 RT Gene 147:231-235 (1994).
 RN [3] _
 RP SEQUENCE FROM N.A.
 RC STRAIN=White leghorn;
 MEDLINE#=87146391; PubMed=3029691;
 RX
 RA Mezquita J., Pau M., Mezquita C.;
 RT "A novel carbonic anhydrase II mRNA isolated from mature chicken
 testis displays a TATA box and other promoter sequences in a leader
 5' untranslated region not present in somatic tissues.";
 RT Gene 147:231-235 (1994).

"The chicken carbonic anhydrase II gene: evidence for a recent shift in intron position." ; Nucleic Acids Res. 15:753-770 (1987). [3]

SEQUENCE OF 4-259 FROM N.A. ; SEQUENCE OF 221-259 FROM N.A.

STRAIN=White Leghorn; TISSUE=Retina;
MEDLINE=81133522; PubMed=3102231;

Rogers J.H.;
"Sequence of carbonic anhydrase II cDNA from chick retina." ; Eur. J. Biochem. 162:119-122 (1987). [4]

SEQUENCE OF 7-86 FROM N.A.

STRAIN=Retina;
MEDLINE=90192090; PubMed=1969140;

Godbout R.; Anderson R.; Upton C.; Day R.;
Utilization of the second polyadenylation signal at the 3' end of
the chicken carbonic anhydrase II gene." ;
Nucleic Acids Res. 18:1049-1049 (1990). [5]

SEQUENCE OF 7-86 FROM N.A.

STRAIN=Retina;
MEDLINE=88255154; PubMed=6131256;

Yoshihara C.M.; Federpiel M.; Dodgson J.B.;
"Isolation of the chicken carbonic anhydrase II gene." ;
Ann. N.Y. Acad. Sci. 429:332-334 (1984).

FUNCTION: Reversible hydration of carbon dioxide.
Catalytic activity: H₂CO(3) = CO(2) + H₂O.
COFACTOR: Zinc.
SUBCELLULAR LOCATION: Cytoplasmic.
SIMILARITY: Belongs to the eukaryotic-type carbonic anhydrase family.

SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce> send an email to license@isb-sib.ch).

| | | | |
|----|---------|-------------|---------|
| L; | Z14957; | CAA78681.1; | -; |
| L; | X12339; | CAA31175.1; | -; |
| L; | X06000; | CAA29417.1; | -; |
| L; | X06001; | CAA29417.1; | JOINED. |
| L; | X06002; | CAA29417.1; | JOINED. |
| L; | X06003; | CAA29417.1; | JOINED. |
| L; | X06004; | CAA29417.1; | JOINED. |
| L; | X06005; | CAA29417.1; | JOINED. |
| L; | X04810; | CAA28501.1; | -; |
| L; | X17378; | CAA32550.1; | -; |
| L; | M25943; | AAA48646.1; | -; |
| L; | PC2280; | JC2580. | |
| P; | P00918; | ICLM. | |

| Db | 4 | WGYDSSINGPAWKHEHFTIANGRQSPIAISTGARYPFAKPKPSFSDAGTAKAIVNNGH | 63 |
|---|-----------------------------------|---|---------|
| Qy | 66 | SENVDFDTENKSVLQRGPGLTGSYRLQEVHLIWGADDHGSEHILUDGYSYAAELHVYHN | 125 |
| Db | 64 | SENVEFDSSDKSVLQGGALGVYRLVQFHITHWGSCBQGSEETVTDGKYDAELHVYHN | 123 |
| Qy | 126 | SDKPYSVEAAHEPDGHAVLGFQLQGPNPSQOKTPDLSIREKGQOTRFNTFDLISL | 185 |
| Db | 124 | V-KYGPKAELKHFDGIAVVGFMKGVNKAPEHQKVDAINSQTKGQASFNTFDPTGL | 182 |
| Qy | 186 | LPPSWDYWTWYPGSLTVPPLESWTVWIKQPINISSQOLAKERPLSLCTAEGE | 237 |
| Db | 183 | LPPCRDWTWYPGSLTTPLHBCVTHWLKEPTVSSEOMCKRGLCFSAEAE | 234 |
| RESULT 14 | | | |
| CAHL_SHEEP | | STANDARD ; | |
| ID_CAH1_SHEEP | | PRT ; | 260 AA. |
| AC_P48282; | | | |
| DT 01-FEB-1996 | (Rel. 33, Created) | | |
| DT 01-FEB-1996 | (Rel. 33, Last sequence update) | | |
| DT 10-OCT-2003 | (Rel. 42, Last annotation update) | | |
| DE Carbonic anhydrase I (EC 4.2.1.1) (Carbonic anhydratase I) (CA-I). | | | |
| GN CA1. | | | |
| OS Ovis aries (Sheep). | | | |
| OC Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Buteleostomi; | | | |
| Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea; | | | |
| Parvorder: Cervinae; Order: Cervidae. | | | |

| Sequence Comparison Results | | | | | | | | |
|-----------------------------|--|------------------------------------|-----------------------------------|------------------------------------|---|----------------|-----------|-------------------------|
| | Query | Match | Score | DB | Length | Gaps | | |
| CC | EMBL; L42178; AAC01634; 1; | - | - | - | - | - | | |
| CC | DR HSSP; P00915; 1BZN. | InterPro; IPR001148; Bulk COanhds. | PFam; PF00194; carb_anhydrase; 1. | prodDom; PD000865; Euk_Coanhds; 1. | PROSITE; PS00162; EUK_CO2_ANHYDRASE; 1. | KW | | |
| CC | DR | lysine; Zinc; Metal-binding. | INTT_MET | 0 | 0 | BY SIMILARITY. | | |
| CC | FT METAL | 94 | 94 | ZINC (CATALYTIC). | FT METAL | 96 | 96 | ZINC (CATALYTIC). |
| CC | FT METAL | 119 | 119 | ZINC (CATALYTIC). | SEQUENCE | 260 AA; | 28766 MW; | CB96A0875CB041CE CRC64; |
| CC | SEQUENCE | 60.1%; Score 775; | DB 1; | Length 260; | | | | |
| CC | Best Local Similarity | 60.3%; | Pred. No. 1-8e-60; | | | | | |
| CC | Matches 140; Conservative | 39; | Mismatches 53; | Indels 0; | | | | |
| CC | | | | | | Gaps | | |
| Qy | 6 WGYREHNGPIHWKEEFFPIADGDOQSPIEIKTKEYKDSSLRPLASIKYDPSAKIISNSGH | | | | | 65 | | |
| Db | 5 WGYDGENGPHEWCKLHPIANGNNQSPIDKTSSETKRDPSLKPLSISYNPAKIEIVNVGH | | | | | 64 | | |

| | | | |
|-----------|---|---|--------------|
| Qy | 66 | SENVDFDDTENKSVLRGGPGLTGSPYLQVHLWGGSADDHGSEHIVYDGVSYAELHHVHN | 125 |
| Qy | 65 | SEHNFEDSDNRSTVKGGPLPESIRLQQFHHGSGTDCSEHLLDGATFAELLHVHN | 124 |
| Db | 126 | SDKYPSVFAAHEDPLGLAVLGFLQIGEPNSQLOKTDTSIKKGKQIRFTNFDLISI | 185 |
| Db | 125 | SAKYPSFADASQGLVYGVLMVGQANPNLQCVLDALKTVKTKNNKKAPFTNFDPSVL | 184 |
| Qy | 186 | LPPSWDYWTYPPGSLTPPLLESVTWIVLKPONISSQQLAKFRSLICTAEGE | 237 |
| Db | 185 | LPSPCPDWAYFGSLTHPPHLSESVTWILFETISVSQAEQAFRSLLANAEGD | 236 |
| RESULT 15 | | | |
| | CAH1_MOUSE | SEQUENCE FROM N.A. | |
| | ID CAH1_MOUSE | STANDARD ; | PRT, 260 AA. |
| | AC P13634 | PRD0C84; | |
| | DT 01-JAN-1990 | (Rel. 13, Created) | |
| | DT 01-FEB-1996 | (Rel. 33, Last sequence update) | |
| | DT 10-OCT-2003 | (Rel. 42, Last annotation update) | |
| | DR Carbonic anhydrase I (EC 4.2.1.1) | (Carbonate dehydrase I) (CA-I). | |
| | GN CAI OR CARI. | | |
| | OS Mus musculus (Mouse). | | |
| | OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; | | |
| | OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. | | |
| | NCBI_TaxID=10090; | | |
| | RN [1] | | |
| | RP SEQUENCE FROM N.A. | | |
| | RX MEDLINE=87169766; PubMed=3104601; | | |
| | RA Fraser P.J., Curris P.J.; | | |
| | RA RT "Molecular evolution of the carbonic anhydrase genes: calculation of divergence time for mouse carbonic anhydrase I and II."; | | |
| | J. Mol. Evol. 23:294-299 (1986). | | |
| | [2] | | |
| | RN [2] | | |
| | RP SEQUENCE FROM N.A. | | |
| | RC TISSUE=Brain; | | |
| | RX MEDLINE=90014784; PubMed=2571923; | | |
| | RA Fraser P.J., Cummings P., Curtis P.J.; | | |
| | RA RT "The mouse carbonic anhydrase I gene contains two tissue-specific promoters."; | | |
| | RA RT Mol. Cell. Biol. 9:3308-3313 (1989). | | |
| | [3] | | |
| | RN [3] | | |
| | RP SEQUENCE FROM N.A. | | |
| | RC STRAIN=C57BL/6J; TISSUE=Spleen; | | |
| | RX MEDLINE=2108560; PubMed=11217851; | | |
| | RA Kawai J., Shinagawa A., Shibusawa K., Yoshino M., Itoh M., Ishii Y., | | |
| | RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S., | | |
| | RA Aizawa K., Izawa M., Nishii K., Kiyosawa H., Yamana T., R., | | |
| | RA Saito T., Okazaki Y., Gojobori T., Bono H., Katsukawa T., Saito R., | | |
| | RA Kadota K., Matsusada H.A., Ashburner M., Batzalov S., Casavant T., | | |
| | RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H., | | |
| | RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J., | | |
| | RA Schramm L.M., Straubli F., Suzuki R., Tomita M., Wagner L., Washio T., | | |
| | RA Sakai K., Okido T., Furuno M., Aono H., Baldacci R., Barth G., | | |
| | RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F., | | |
| | RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M., | | |
| | RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H., | | |
| | RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P., | | |
| | RA Nordne P., Ring B., Ringwald M., Rodriguez I., Sakanoto N., | | |
| | RA Sasaki H., Sato K., Schoenbach C., Shibata Y., Storch K.-F., | | |
| | RA Suzuki H., Toyo-oka K., Wang K.H., Weitz C., Whittaker C., Wilming L., | | |
| | RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohetsuki S., | | |
| | RA Hyashizaki Y.; | | |
| | RA "Functional annotation of a full-length mouse cDNA collection."; | | |
| | RL Nature 409:685-690 (2001). | | |
| | [4] | | |
| | RP SEQUENCE FROM N.A. | | |
| | RC TISSUE=Colon; | | |
| | RX MEDLINE=22188257; PubMed=12477932; | | |
| | RA Strasberg R.L., Feingold E.A., Grouse L.H., Derge J.G., | | |
| | RA Klausner R.D., Collins F.S., Wagner L., Shemmen C.M., Schuler G.D., | | |
| | RA Altschul S.F., Zeeberg B., Buetow K.H., Scheuer C.F., Bhat N.K., | | |
| | RA Hopkins R.F., Jordan H., Moore T., Max S.T., Wang J., Heisler P. | | |
| | RA "Functional annotation of a full-length mouse cDNA collection."; | | |
| | RL Nature 409:685-690 (2001). | | |
| | [4] | | |

185 LPSSLDWWTYFGSLTLHPPPLMESVTWICKDSISLSSPEQLAQRLGILSSAEGESA. 238

Search completed: September 9, 2004, 15:11:56
Job time : 29 secs

Copyright (C) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 9, 2004, 15:06:00 ; Search time 115 Seconds
(without alignments)
663.960 Million cell updates/sec

Title: US-10-069-434-1
Perfect score: 1289
Sequence: 1 MSRLSWGYREHNGPHWKEF.....QLAKFRSILCTAEGEAAFL 242

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 101041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing First 45 summaries

Database : SPTRMBL_25:
1: sp_archaea:
2: sp_bacteria:
3: sp_fungi:
4: sp_human:
5: sp_invertebrate:
6: sp_mammal:
7: sp_other:
8: sp_organelle:
9: sp_phage:
10: sp_plant:
11: sp_rabbit:
12: sp_virus:
13: sp_vertebrate:
14: sp_unclassified:
15: sp_rvirus:
16: sp_bacteriopl:
17: sp_archeap:

ALIGNMENTS

| | Score | Query | Match | Length | DB | ID | Description |
|----|-------|-------|-------|--------|--------|------------|---------------------|
| 1 | 800 | 62.1 | 261 | 13 | Q8JG56 | | Q8ij956 lepisosteus |
| 2 | 794.5 | 61.6 | 260 | 6 | Q86577 | bos taurus | Q8T2K6 oncorhynchus |
| 3 | 775.5 | 60.2 | 260 | 13 | Q7T2K6 | | Q7T2K6 oncorhynchus |
| 4 | 751.5 | 58.3 | 260 | 11 | Q7TP61 | | Q7TP61 mus musculus |
| 5 | 745.5 | 57.8 | 260 | 13 | Q8AVG8 | | Q8avg8 xenopus lae |
| 6 | 721.5 | 56.0 | 260 | 13 | Q7ZYU6 | | Q7zyu6 xenopus lae |
| 7 | 675 | 52.4 | 264 | 11 | Q811X4 | | Q811X4 mus musculus |
| 8 | 666 | 51.7 | 306 | 13 | Q7ZUE2 | | Q7zue2 brachydanio |
| 9 | 551 | 42.7 | 208 | 4 | Q86YU0 | | Q86YU0 homo sapiens |
| 10 | 548.5 | 42.6 | 261 | 5 | Q9XZG6 | | Q9xZG6 anchopleura |
| 11 | 546.5 | 42.4 | 270 | 5 | Q9V396 | | Q9v396 dirosophila |
| 12 | 546.5 | 42.4 | 192 | 11 | Q9DCT3 | | Q9dct3 mus musculus |
| 13 | 487.5 | 37.8 | 243 | 5 | Q8MPH8 | | Q8mpH8 riftia pach |
| 14 | 486.5 | 33.9 | 344 | 11 | Q8K2U1 | | Q93K87 platichthys |
| 15 | 406 | 31.5 | 259 | 13 | Q93K87 | | Q66218 cowpox virus |
| 16 | 393 | 30.5 | 304 | 12 | Q66218 | | |

| | Score | Query | Match | Length | DB | ID | Description |
|----|-------|-------|-------|--------|--------|----|-------------------------------|
| 1 | 392 | 30.4 | 304 | 12 | Q8QMV3 | | Q8QMV3 cowpox virus |
| 2 | 392 | 30.4 | 304 | 12 | Q9JFA1 | | Q9jfa1 vaccinia virus |
| 3 | 19 | 30.2 | 304 | 12 | Q90197 | | Q90197 monkeypox virus |
| 4 | 20 | 30.1 | 304 | 12 | Q80DX3 | | Q80dx3 cowpox virus |
| 5 | 21 | 29.9 | 304 | 12 | Q8JLB8 | | Q8Jlb8 ectromelia virus |
| 6 | 22 | 29.9 | 304 | 12 | Q8V4Y0 | | Q8v4y0 monkeypox virus |
| 7 | 23 | 29.9 | 304 | 12 | Q8V2R1 | | Q8v2r1 canarypox virus |
| 8 | 24 | 29.9 | 304 | 12 | Q66281 | | Q66281 camelpox virus |
| 9 | 25 | 29.7 | 304 | 12 | Q67211 | | Q67211 vaccinia virus |
| 10 | 26 | 29.6 | 327 | 5 | Q9W316 | | Q9w316 drosophila pseudopicta |
| 11 | 27 | 29.2 | 110 | 13 | Q7SXQ8 | | Q7sxq8 ectromelia virus |
| 12 | 28 | 29.1 | 303 | 5 | Q8A439 | | Q8a439 shopo fibro |
| 13 | 29 | 29.1 | 285 | 12 | Q9Q8Z4 | | Q9q8z4 canis familiaris |
| 14 | 30 | 28.5 | 320 | 6 | Q865C0 | | Q865c0 fasciatus |
| 15 | 31 | 28.3 | 286 | 12 | Q9QBL7 | | Q9qbl7 myzoma virum |
| 16 | 32 | 27.5 | 317 | 11 | Q7TNG9 | | Q7tn9g mus musculus |
| 17 | 33 | 27.5 | 325 | 11 | Q80YB7 | | Q80yb7 mus musculus |
| 18 | 34 | 27.1 | 335 | 5 | Q9VTU8 | | Q9vtu8 drosophila |
| 19 | 35 | 26.3 | 328 | 6 | Q9N085 | | Q9n085 macaca fasciata |
| 20 | 36 | 26.2 | 250 | 5 | Q9W3C8 | | Q9w3c8 pastorella |
| 21 | 37 | 26.2 | 252 | 16 | Q9CJ76 | | Q9cj76 drosophila |
| 22 | 38 | 23.5 | 311 | 5 | Q9VB76 | | Q9vb76 ratetus norvegicus |
| 23 | 39 | 23.5 | 328 | 11 | Q811X3 | | Q811x3 rat |
| 24 | 40 | 23.2 | 312 | 13 | Q7SYW3 | | Q7syw3 xenopus laevis |
| 25 | 41 | 23.0 | 306 | 5 | Q866X7 | | Q866x7 bos taurus |
| 26 | 42 | 23.0 | 306 | 7 | Q866X6 | | Q866x6 sus scrofa |
| 27 | 43 | 23.5 | 303 | 5 | Q9i908 | | Q9i908 xenopus laevis |
| 28 | 44 | 23.5 | 227 | 13 | Q9i909 | | Q9i909 xenopus laevis |
| 29 | 45 | 23.3 | 275 | 16 | Q9KFW1 | | Q9kfw1 bacillus haemolyticus |

| | Score | Query | Match | Length | DB | ID | Description |
|----|-------|-------|-------|--------|--------|----|-------------------------------|
| 1 | 17 | 30.4 | 304 | 12 | Q8QMV3 | | Q8QMV3 cowpox virus |
| 2 | 18 | 30.4 | 304 | 12 | Q9JFA1 | | Q9jfa1 vaccinia virus |
| 3 | 19 | 30.2 | 304 | 12 | Q90197 | | Q90197 monkeypox virus |
| 4 | 20 | 30.1 | 304 | 12 | Q80DX3 | | Q80dx3 cowpox virus |
| 5 | 21 | 29.9 | 304 | 12 | Q8JLB8 | | Q8Jlb8 ectromelia virus |
| 6 | 22 | 29.9 | 304 | 12 | Q8V4Y0 | | Q8v4y0 monkeypox virus |
| 7 | 23 | 29.9 | 304 | 12 | Q8V2R1 | | Q8v2r1 canarypox virus |
| 8 | 24 | 29.9 | 304 | 12 | Q66281 | | Q66281 camelpox virus |
| 9 | 25 | 29.7 | 304 | 12 | Q67211 | | Q67211 vaccinia virus |
| 10 | 26 | 29.6 | 327 | 5 | Q9W316 | | Q9w316 drosophila pseudopicta |
| 11 | 27 | 29.2 | 110 | 13 | Q7SXQ8 | | Q7sxq8 ectromelia virus |
| 12 | 28 | 29.1 | 303 | 5 | Q8A439 | | Q8a439 shopo fibro |
| 13 | 29 | 29.1 | 285 | 12 | Q9Q8Z4 | | Q9q8z4 canis familiaris |
| 14 | 30 | 28.5 | 320 | 6 | Q865C0 | | Q865c0 fasciatus |
| 15 | 31 | 28.3 | 286 | 12 | Q9CJ76 | | Q9cj76 pastorella |
| 16 | 32 | 27.5 | 317 | 11 | Q9VB76 | | Q9vb76 drosophila |
| 17 | 33 | 27.5 | 325 | 11 | Q811X3 | | Q811x3 rat |
| 18 | 34 | 27.1 | 335 | 5 | Q9VTU8 | | Q9vtu8 drosophila |
| 19 | 35 | 26.3 | 328 | 6 | Q9N085 | | Q9n085 macaca fasciata |
| 20 | 36 | 26.2 | 250 | 5 | Q9W3C8 | | Q9w3c8 pastorella |
| 21 | 37 | 26.2 | 252 | 16 | Q9CJ76 | | Q9cj76 drosophila |
| 22 | 38 | 23.5 | 311 | 5 | Q9VB76 | | Q9vb76 ratetus norvegicus |
| 23 | 39 | 23.5 | 328 | 11 | Q811X3 | | Q811x3 rat |
| 24 | 40 | 23.2 | 312 | 13 | Q7SYW3 | | Q7syw3 xenopus laevis |
| 25 | 41 | 23.0 | 306 | 5 | Q866X7 | | Q866x7 bos taurus |
| 26 | 42 | 23.0 | 306 | 7 | Q866X6 | | Q866x6 sus scrofa |
| 27 | 43 | 23.5 | 303 | 5 | Q9i908 | | Q9i908 xenopus laevis |
| 28 | 44 | 23.5 | 227 | 13 | Q9i909 | | Q9i909 xenopus laevis |
| 29 | 45 | 23.3 | 275 | 16 | Q9KFW1 | | Q9kfw1 bacillus haemolyticus |

| | Score | Query | Match | Length | DB | ID | Description |
|----|-------|-------|-------|--------|--------|----|-------------------------------|
| 1 | 17 | 30.4 | 304 | 12 | Q8QMV3 | | Q8QMV3 cowpox virus |
| 2 | 18 | 30.4 | 304 | 12 | Q9JFA1 | | Q9jfa1 vaccinia virus |
| 3 | 19 | 30.2 | 304 | 12 | Q90197 | | Q90197 monkeypox virus |
| 4 | 20 | 30.1 | 304 | 12 | Q80DX3 | | Q80dx3 cowpox virus |
| 5 | 21 | 29.9 | 304 | 12 | Q8JLB8 | | Q8Jlb8 ectromelia virus |
| 6 | 22 | 29.9 | 304 | 12 | Q8V4Y0 | | Q8v4y0 monkeypox virus |
| 7 | 23 | 29.9 | 304 | 12 | Q8V2R1 | | Q8v2r1 canarypox virus |
| 8 | 24 | 29.9 | 304 | 12 | Q66281 | | Q66281 camelpox virus |
| 9 | 25 | 29.7 | 304 | 12 | Q67211 | | Q67211 vaccinia virus |
| 10 | 26 | 29.6 | 327 | 5 | Q9W316 | | Q9w316 drosophila pseudopicta |
| 11 | 27 | 29.2 | 110 | 13 | Q7SXQ8 | | Q7sxq8 ectromelia virus |
| 12 | 28 | 29.1 | 303 | 5 | Q8A439 | | Q8a439 shopo fibro |
| 13 | 29 | 29.1 | 285 | 12 | Q9Q8Z4 | | Q9q8z4 canis familiaris |
| 14 | 30 | 28.5 | 320 | 6 | Q865C0 | | Q865c0 fasciatus |
| 15 | 31 | 28.3 | 286 | 12 | Q9CJ76 | | Q9cj76 pastorella |
| 16 | 32 | 27.8 | 317 | 5 | Q9VFH1 | | Q9vfh1 myzoma virum |
| 17 | 33 | 27.8 | 325 | 11 | Q8MPH8 | | Q8mph8 riftia pachynema |
| 18 | 34 | 27.8 | 334 | 11 | Q8K2U1 | | Q8k2u1 mus musculus |
| 19 | 35 | 27.8 | 334 | 11 | Q93K87 | | Q93k87 platichthys flesus |
| 20 | 36 | 27.8 | 334 | 11 | Q9i908 | | Q9i908 xenopus laevis |
| 21 | 37 | 27.8 | 334 | 11 | Q9i909 | | Q9i909 xenopus laevis |
| 22 | 38 | 27.8 | 334 | 11 | Q9KFH1 | | Q9kfh1 bacillus haemolyticus |
| 23 | 39 | 27.8 | 334 | 11 | Q9V396 | | Q9v396 cowpox virus |
| 24 | 40 | 27.8 | 334 | 11 | Q9V4Y0 | | Q9v4y0 monkeypox virus |
| 25 | 41 | 27.8 | 334 | 11 | Q9V2R1 | | Q9v2r1 canarypox virus |
| 26 | 42 | 27.8 | 334 | 11 | Q9VJG6 | | Q9vjg6 ectromelia virus |
| 27 | 43 | 27.8 | 334 | 11 | Q9VJF1 | | Q9vjf1 vaccinia virus |
| 28 | 44 | 27.8 | 334 | 11 | Q9VJL8 | | Q9vjl8 ectromelia virus |
| 29 | 45 | 27.8 | 334 | 11 | Q9VJL9 | | Q9vjl9 ectromelia virus |
| 30 | 46 | 27.8 | 334 | 11 | Q9VJL9 | | Q9vjl9 ectromelia virus |
| 31 | 47 | 27.8 | 334 | 11 | Q9VJL9 | | Q9vjl9 ectromelia virus |
| 32 | 48 | 27.8 | 334 | 11 | Q9VJL9 | | Q9vjl9 ectromelia virus |
| 33 | 49 | 27.8 | 334 | 11 | Q9VJL9 | | Q9vjl9 ectromelia virus |
| 34 | 50 | 27.8 | 334 | 11 | Q9VJL9 | | Q9vjl9 ectromelia virus |
| 35 | 51 | 27.8 | 334 | 11 | Q9VJL9 | | Q9vjl9 ectromelia virus |
| 36 | 52 | 27.8 | 334 | 11 | Q9VJL9 | | Q9vjl9 ectromelia virus |
| 37 | 53 | 27.8 | 334 | 11 | Q9VJL9 | | Q9vjl9 ectromelia virus |
| 38 | 54 | 27.8 | 334 | 11 | Q9VJL9 | | Q9vjl9 ectromelia virus |
| 39 | 55 | 27.8 | 334 | 11 | Q9VJL9 | | Q9vjl9 ectromelia virus |
| 40 | 56 | 27.8 | 334 | 11 | Q9VJL9 | | Q9vjl9 ectromelia virus |
| 41 | 57 | 27.8 | 334 | 11 | Q9VJL9 | | Q9vjl9 ectromelia virus |
| 42 | 58 | 27.8 | 334 | 11 | Q9VJL9 | | Q9vjl9 ectromelia virus |
| 43 | 59 | 27.8 | 334 | 11 | Q9VJL9 | | Q9vjl9 ectromelia virus |
| 44 | 60 | 27.8 | 334 | 11 | Q9VJL9 | | Q9vjl9 ectromelia virus |
| 45 | 61 | 27.8 | 334 | 11 | Q9VJL9 | | Q9vjl9 ectromelia virus |
| 46 | 62 | 27.8 | 334 | 11 | Q9VJL9 | | Q9vjl9 ectromelia virus |
| 47 | 63 | 27.8 | 334 | 11 | Q9VJL9 | | Q9vjl9 ectromelia virus |
| 48 | 64 | 27.8 | 334 | 11 | Q9VJL9 | | Q9vjl9 ectromelia virus |
| 49 | 65 | 27.8 | 334 | 11 | Q9VJL9 | | Q9vjl9 ectromelia virus |
| 50 | 66 | 27.8 | 334 | 11 | Q9VJL9 | | Q9vjl9 ectromelia virus |
| 51 | 67 | 27.8 | 334 | 11 | Q9VJL9 | | Q9vjl9 ectromelia virus |
| 52 | 68 | 27.8 | 334 | 11 | Q9VJL9 | | Q9vjl9 ectromelia virus |
| 53 | 69 | 27.8 | 334 | 11 | Q9VJL9 | | Q9vjl9 ectromelia virus |
| 54 | 70 | 27.8 | 334 | 11 | Q9VJL9 | | Q9vjl9 ectromelia virus |
| 55 | 71 | 27.8 | 334 | 11 | Q9VJL9 | | Q9vjl9 ectromelia virus |
| 56 | 72 | 27.8 | 334 | 11 | Q9VJL9 | | Q9vjl9 ectromelia virus |
| 57 | 73 | 27.8 | 334 | 11 | Q9VJL9 | | Q9vjl9 ectromelia virus |
| 58 | 74 | 27.8 | 334 | 11 | Q9VJL9 | | Q9vjl9 ectromelia virus |
| 59 | 75 | 27.8 | 334 | 11 | Q9VJL9 | | Q9vjl9 ectromelia virus |
| 60 | | | | | | | |

| | | | | | |
|----------|--|---|------|--|--|
| 6 | HSFQVDFADENDSSTLQQGPISGVYRLRQFHGASDERSERSEHTVGGVYAAEGLHLVHW | 123 | OC | Actinopterygii; Neopterigii; Teleostei; Buteleostei; | |
| Db | | | OC | Protacanthopterygii; Salmoniformes; Salmonidae; Oncorhynchus. | |
| QY | 125 NSDPSWVTPGSLTVPPLIESVTWVLKOPINISSQQLAKERSLCTAEGAA | 184 | RN | SEQUENCE FROM N.A. | |
| Db | 124 NAGKXASFGDRAKAPDGLAVGVFLKGIGAASPNLQKVYDALLDAIKTKGKQTDPFQDPKI | 183 | RP | Esborgh A.J.; Lund S.G.; Tufts B.L.; | |
| | | | RT | "The Evolution of Red Blood Cell Carbonic Anhydrase: Properties of RBC CA in a Teleost Fish, <i>Oncorhynchus mykiss</i> ," | |
| | | | RT | Submitted (MAY-2003) to the ENSEMBL GenBank/DBJ databases. | |
| Db | 185 LPPSWDWTYTPGSLTVPPLIESVTWVLKOPINISSQQLAKERSLCTAEGAA | 239 | RL | EMBL; AY307022; AAPJ3748.1; - | |
| | | | DR | SEQUENCE 260 AA; 28613 MW; 5C928B72B922A52 CRC64; | |
| | | | SQ | | |
| RESULT 2 | Q865Y7 | PRELIMINARY; | PRT; | 260 AA. | Query Match 60.2%; Score 775.5; DB 13; Length 260; |
| | ID Q865Y7 | | | | Best Local Similarity 60.7%; Pred. No. 4.1e-61; |
| | AC Q865Y7 | | | | Mismatches 37; Indels 1; Gaps 1 |
| | DT 01-JUN-2003 | (TREMBLrel. 24, Created) | | | Matches 142; Conservative 37; Mismatches 54; |
| | DT 01-OCT-2003 | (TREMBLrel. 24, Last sequence update) | | | |
| | DT 01-OCT-2003 | (TREMBLrel. 25, Last annotation update) | | | |
| | DE | Carbonic anhydrase II (EC 4.2.1.1). | | | |
| | OS Bos taurus (Bovine). | | | | |
| | OSBaryta; | | | | |
| | Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; | | | | |
| | Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Bovidae; | | | | |
| | Bovidae; Bovinae; Bos. | | | | |
| | OX NCBI_TaxID=9913; | | | | |
| | RN [1]; | | | | |
| | RP TISSUE-Bone marrow; | | | | |
| | RC Daigle R., Castro I., Desrochers M., Charest P.-M.; | | | | |
| | RT "Full length cDNA of Bovine Carbonic Anhydrase II," | | | | |
| | RT submitted (FEB-2003) to the EMBL/GenBank/DBJ databases. | | | | |
| | DR AY240020; AA085140.1; | | | | |
| | DR GO; GO:0004089; R:carbonate dehydratase activity; IEA. | | | | |
| | DR GO; GO:0006829; R:lyase activity; IEA. | | | | |
| | DR GO; GO:0008270; R:zinc ion binding; IEA. | | | | |
| | DR GO; GO:0006730; R:one-carbon compound metabolism; IEA. | | | | |
| | DR InterPro01148; Buk Coanh. | | | | |
| | DR Pfam; PF00194; carb anhydrase_1. | | | | |
| | DR ProDom; PDD000865; Euk_Coanh; 1. | | | | |
| | DR PROSITE; PS00162; EUK_CO2_ANHYDRASE; 1. | | | | |
| | KW Lyase. | | | | |
| | SQ SEQUENCE 260 AA; 29114 MW; A89412G23FFD19A7 CRC64; | | | | |
| RESULT 2 | Q865Y7 | PRELIMINARY; | PRT; | 260 AA. | Query Match 61.6%; Score 794.5; DB 6; Length 260; |
| | ID Q865Y7 | | | | Best Local Similarity 62.5%; Pred. No. 8.2e-63; |
| | AC Q865Y7 | | | | Mismatches 39; Indels 1; Gaps 1; |
| | DT 01-OCT-2003 | (TREMBLrel. 25, Last sequence update) | | | |
| | DT 01-OCT-2003 | (TREMBLrel. 25, Last annotation update) | | | |
| | DE | Hypothetical protein. | | | |
| | OS Mus musculus (Mouse). | | | | |
| | OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; | | | | |
| | OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. | | | | |
| | NCBI_TaxID=10090; | | | | |
| | RN | | | | |
| | RP SEQUENCE FROM N.A. | | | | |
| | RC STRAIN=FVB/N; TISSUE=Colon; | | | | |
| | RX MEDLINE=2238257; PubMed=12477932; | | | | |
| | RA Strausberg R.L.; Feingold E.A.; Grouse L.H.; Derge J.G., | | | | |
| | RA Klausner R.D.; Collins F.S.; Wagner L.; Shanmen C.M.; Schuler G.D., | | | | |
| | RA Alleschul S.F.; Zeeberg B.; Buetow K.H.; Schaefer C.F.; Bhat N.K., | | | | |
| | RA Hopkins R.F.; Jordan H.; Moore T.; Max S.I.; Wang J.; Hsieh F., | | | | |
| | RA Diatchenko L.; Matsunaga K.; Farmer A.A.; Bonaldo M.B.; Bontalio M.F.; Casavant T.L.; Scheetz T.E., | | | | |
| | RA Stapleton M.; Brownstein M.J.; Urdinolain T.B.; Toshiyuki S.; Carrinco P.; Prange C., | | | | |
| | RA Brownstein M.J.; Urdinolain T.B.; Toshiyuki S.; Carrinco P.; Prange C., | | | | |
| | RA Raha S.S.; Loquellano N.A.; Peters G.J.; Abramson R.D.; Mulhall S.J., | | | | |
| | RA Bosak S.A.; McEwan P.J.; McKernan K.J.; Malek J.A.; Gunaratne P.H., | | | | |
| | RA Richards S.; Worley K.; Hale S.; Garcia A.M.; Gay L.J.; Hulyk S.W., | | | | |
| | RA Villalon D.K.; Munay D.M.; Sodergren B.J.; Lu X.; Gibbs R.A., | | | | |
| | RA Fahy J.; Heiton E.; Kettman M.; Madan A.; Rodriguez S.; Sanchez A., | | | | |
| | RA Whalesley R.W.; Touchman J.W.; Green E.D.; Dickson M.C., | | | | |
| | RA Rodriguez A.C.; Grimwood J.; Schmutz J.; Myers R.M.; Butterfield Y.S., | | | | |
| | RA Krzywinski M.I.; Skalska U.; Smalius D.E.; Schnurch A.; Schein J.E., | | | | |
| | RA Jones S.J.; Marra M.A.; Sodergren B.J.; Lu X.; Gibbs R.A., | | | | |
| | RA Fahy J.; Heiton E.; Kettman M.; Madan A.; Rodriguez S.; Sanchez A., | | | | |
| | RA Whalesley R.W.; Touchman J.W.; Green E.D.; Dickson M.C., | | | | |
| | RA Rodriguez A.C.; Grimwood J.; Schmutz J.; Myers R.M.; Butterfield Y.S., | | | | |
| | RA Strain=FVB/N; Tissue=Colon; | | | | |
| | RA Strausberg R. | | | | |

RL Submitted (JUL-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC05291; AAH5291.1; -.
KW Hypothetical protein.
SQ SEQUENCE 260 AA; 29041 MW; 43FB1AC12238C5E4 CRC64;

RESULT 6

Query Match 58.3%; Score 751.5; DB 11; Length 260;
Best Local Similarity 58.4%; Pred. No. 5; e-59;
Matches 136; Conservative 42; Mismatches 54; Indels 1; Gaps 1;
Qy 6 WGYREINGPILWKEFFPPIADGQQSPPIERIKTEKVYDSSURPLSIKYDPSAKIISNSGH 65
Db 5 WGSKRSKNGPENTWKDFPIANGDQRSPVDDTATAHHDPALQLLISYDKAASKSITVNGH 64
Qy 66 SENVDFDTENKSVLRGGLPTGSYRLRQVHLHWGSADDHGSEHIVDGVSAYAELHVHWN 125
Db 65 SENVFEDDSQDAVLRQGPSSPSYRLIQFHFWGSSQGSHTVNKKYAAELHVHWN 124

Qy 126 SDKYP5FVEAAHEPDGLAVLGVLQIGEPNSLOQKLTDTLSIKEKGKQTRTNFELL 185
Db 125 T-KYDFGKAVQQPDPGLAVLGFLKGPAQSLQKYLEAHSIKTKGKRAAFANFOPCSL 183
Qy 186 LPPSWDYWTYPCSSLITPPPLLESVTWVYLQKPINNISQOLAKERSLIICTAEGA 238
Db 184 LPGNLQDWTYPCSSLITPPPLECYQSPINIVPAEAMYDQHUKPISKYDSTAKVILNG 236

RESULT 5

Q8AVG8 PRELIMINARY; PRT; 260 AA.
ID Q8AVG8; PRELIMINARY; PRT; 260 AA.
AC 08AVG8; PRELIMINARY; PRT; 260 AA.
DT 01-MAR-2003 (TREMBLrel. 23, Created)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE Similar to carbonic anhydrase II.
OS Xenopus laevis (African clawed frog).
OC Amphibia; Batrachia; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Xenopoda; Xanopidae; Pipidae;
OC Xenopus.
OX NCBI_TaxID=8355;
RN [1] -
RP SEQUENCE FROM N.A.
RC TISSUE=Embryo;
RA Klein S.; Strausberg R.;
RL Submitted (DEC-2002) to the EMBL/GenBank/DBJ databases.
EMBL; BC041213; AAH1213.1; -.
GO; GO:0004089; F:carbonate dehydratase activity; IEA.
GO; GO:0008270; F:zinc ion binding; IEA.
GO; GO:0006730; F:one-carbon compound metabolism; IEA.
InterPro; IPR001148; Euk_Coanhd.
Pfam; PF00194; carb_anhydrase_1.
ProDom; PD000855; Euk_Coanhd; 1.
DR PROSITE; PS000855; Euk_Coanhd; 1.
DR PROSITE; PS00162; Euk_Coanhd; 1.
DR PROSITE; PS000855; Euk_Coanhd; 1.
SQ SEQUENCE 260 AA; 29004 MW; 2C7B5C072C022DA CRC64;

Query Match 56.0%; Score 721.5; DB 13; Length 260;

Best Local Similarity 57.3%; Pred. No. 2; e-56;
Matches 134; Conservative 34; Mismatches 64; Indels 1; Gaps 1;

Qy 5 SWGYREHNGPILWKEFFPPIADGQQSPPIERIKTEKVYDSSURPLSIKYDPSAKIISNSGH 64
Db 4 AWGYGDNGPSTWHHAFLPLKGEYQSPINIVPAEAMYDQHUKPISKYDSTAKVILNG 63
Qy 65 HSFNVDFFDTENKSVLRGGLPTGSYRLQVHLHWGSADDHGSEHIVDGVSAYAELHVHW 124
Db 64 HAFNVFEDDSNRSVSLGGALLESEPYRLQKFHFHNSCEGHGSEHTYVGRCAEELHVHW 123
Qy 125 NSDKYPSFVEAAHEPDGLAVLGVLQIGEPNSLOKLTDTLSIKEKGKQTRTNFELL 184
Db 124 NT-KYGSMAEAVKHCDGLAVGVFLKGPAQSLQKYLEAHSIKTKGKRAAFANFOPCSL 182
Qy 185 LPPSWDYWTYPCSSLITPPPLLESVTWVYLQKPINNISQOLAKERSLIICTAEGA 237
Db 183 LPPNSIDFWTYKGSLLTPPLIQLQVHLWVKEPIAVSKQQLSQRSLFFNAEGD 235

RESULT 7

Q81IX4 PRELIMINARY; PRT; 264 AA.
ID Q81IX4; PRELIMINARY; PRT; 264 AA.
AC 081IX4; PRELIMINARY; PRT; 264 AA.
DT 01-JUN-2003 (TREMBLrel. 24, Created)
DT 01-OCT-2003 (TREMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE Carbonic anhydrase 7.
OS Mus musculus (Mouse).
OC Mammalia; Eutheria; Rodentia; Sciurognath; Muridae; Murinae; Mus.
OX NCBI_TaxID=10900;
RN [1] -
RP SEQUENCE FROM N.A.
RC Chen Y.; Huang C.-H.;
RA "Molecular identification of carbonic anhydrases (CA) and CA-related
RT (CAR) genes.";
RT Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
EMBL; AY075021; AAH78169.1; -.
GO; GO:0004089; F:carbonate dehydratase activity; IEA.
GO; GO:0008270; F:zinc ion binding; IEA.
GO; GO:0006730; F:one-carbon compound metabolism; IEA.
InterPro; IPR001148; Euk_Coanhd.

| | | | | | |
|-----------------------|---|--|--|--|--|
| DR | Pfam; PF00194; carb_anhydride; 1. | | | | |
| DR | ProDom; PD000865; Euk_CoAhd; 1. | | | | |
| DR | PROSITE; PS00162; EUK_CO2_ANHYDRASE; 1. | | | | |
| DR | SEQUENCE; 264 AA; 29915 MW; B58EEB20CB840FA5 CRC64; | | | | |
| Query Match | 52.4%; Score 675; DB 11; Length 264; | | | | |
| Best Local Similarity | 52.2%; Pred. No. 4e-52; | | | | |
| Matches | 121; Conservative 47; Mismatches 64; Indels 0; Gaps 0; | | | | |
| Qy | 6 WGYREHNGPINKKEFPIFPADGQDQSPIEIETKEVYKDSURPLSIKYDPSSAKLISNSGH 65 | | | | |
| Qy | 7 WGYQDQDCPSNWHKLQPAQGDROSPINNISQQAVISPQLQPELFTEACMSLSITNNGH 66 | | | | |
| Db | 66 SPNVDDEDDTENKSVLRGCPPLTSYRQLQVHLWGSADDHGSPHTIDGVSYAAELHVTAN 125 | | | | |
| Qy | 67 SYQVDENDSDDRTVSGGPPLSFPYRLQKHFPTWKEDMGSPSEHTVDKSFSEPLHLVHN 126 | | | | |
| Db | 126 SDKYPSPVEAAHEPDGLFLAVLGVLQFQGPNSQLOKHTDLSIKEKGKQTPTNFDLSSL 185 | | | | |
| Qy | 127 AKCYSTRFEEAAAPDGLAVGVFVLETFGEDEHPSPMNLTAIYMRFKDTKAQFSCENPKCL 186 | | | | |
| Db | 186 LPSPWDWYTYPGSLTVPPILLESVTWIVLKOPINISSQOLAKERSLCTAEGE 237 | | | | |
| Qy | 187 LPSTSRYHTYPGSLTVPPILLESVTWIVLREPISRERMEKERSLLETFSEDD 238 | | | | |
| Db | RESULTS 8 | | | | |
| Q7ZUE2 | PRELIMINARY; PRT; 306 AA. | | | | |
| ID | Q7ZUE2 | | | | |
| AC | 01-JUN-2003 (TREMBLrel. 24, Created) | | | | |
| DT | 01-JUN-2003 (TREMBLrel. 24, Last sequence update) | | | | |
| DR | 01-OCT-2003 (TREMBLrel. 25, Last annotation update) | | | | |
| DE | Similar to carbonic anhydride VII (Fragment). | | | | |
| OS | Brachydanio rerio (Zebrafish) (Danio rerio). | | | | |
| OC | Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; | | | | |
| CC | Actinopterygii; Neopterygii; Ostariophysi; Cypriniformes; | | | | |
| NCBI_TaxID | 7955; | | | | |
| OX | DR Product; PD000865; Euk_CoAhd; 1. | | | | |
| RN | SEQUENCE FROM N.A. | | | | |
| RA | SEQUENCE FROM N.A. | | | | |
| RC | Strasbourg R.; Submitted (MAR-2003) to the EMBL/GenBank/DBBJ databases. | | | | |
| RU | Submitter (MAR-2003) to the EMBL/GenBank/DBBJ databases. | | | | |
| DR | 0004089; AAH49309; AAH49309.1; F:zinc ion binding; IEA. | | | | |
| DR | GO:0008270; F:zinc ion binding; IEA. | | | | |
| DR | GO:0006720; F:one-carbon compound metabolism; IEA. | | | | |
| DR | InterPro; IPR01148; Euk_CoAhd. | | | | |
| DR | PFam; PF00194; carb_anhydride; 1. | | | | |
| DR | ProDom; PD000865; Euk_CoAhd; 1. | | | | |
| DR | PROSITE; PS00162; EUK_CO2_ANHYDRASE; 1. | | | | |
| FT | NON_TER 1 34587 MW; B35078B06366BC05 CRC64; | | | | |
| SQ | SEQUENCE 306 AA; 34587 MW; B35078B06366BC05 CRC64; | | | | |
| Query Match | 51.7%; Score 666; DB 13; Length 306; | | | | |
| Best Local Similarity | 53.0%; Pred. No. 3.2e-51; | | | | |
| Matches | 123; Conservative 40; Mismatches 69; Indels 0; Gaps 0; | | | | |
| Qy | 6 WGYREHNGPINKKEFPIFPADGQDQSPIEIETKEVYKDSURPLSIKYDPSSAKLISNSGH 65 | | | | |
| Qy | 49 WGYGDNGPSAWHKDYPIAEGNRQSPIDIVPSEAFDKLSPIALSYNNCTSLISNSGH 108 | | | | |
| Db | 66 SPNVDDEDDTENKSVLRGCPPLTSYRQLQVHLWGSADDHGSEHIVDGVSYAAELHVTAN 125 | | | | |
| Qy | 109 SVVNEFVDTIERSITGGPLNMTRKQPHFWGSKGCCSEHTVAGKTFVSEPLHLVHN 168 | | | | |
| Db | 126 SDKYPSPVEAAHEPDGLFLAVLGVLQFQGPNSQLOKHTDLSIKEKGKQTPTNFDLSSL 185 | | | | |
| Qy | 169 ANKYKSESEAAVAPGLAVIGIFLETGDERHALQITDALYMREFKGSLAEFKGFENPKCL 228 | | | | |
| Qy | 186 LPPSDWYTYPGSLTVPPILLESVTWIVLKOPINISSQOLAKERSLCTAEGE 237 | | | | |
| Db | RESULTS 9 | | | | |
| Q86YU0 | PRELIMINARY; PRT; 208 AA. | | | | |
| ID | Q86YU0 | | | | |
| AC | Q86YU0 | | | | |
| DT | 01-JUN-2003 (TREMBLrel. 24, Created) | | | | |
| DT | 01-JUN-2003 (TREMBLrel. 24, Last sequence update) | | | | |
| DR | 01-OCT-2003 (TREMBLrel. 25, Last annotation update) | | | | |
| DE | Carbonic anhydrase VII short form. | | | | |
| OS | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; | | | | |
| OC | Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo. | | | | |
| NCBI_TaxID | 9606; | | | | |
| OX | DR Product; PE00194; Carb_anhydride; 1. | | | | |
| RN | SEQUENCE FROM N.A. | | | | |
| RA | Chen Y.; Huang C.-H.; Molecular identification of carbonic anhydrases (CA) and CA-related (CAR) genes.; Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases. | | | | |
| RT | Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases. | | | | |
| RU | EMBL; AY07520; AAL78168.1; -. | | | | |
| DR | GO; GO:0004089; F:carbonate dehydratase activity; IEA. | | | | |
| DR | GO; GO:0008270; F:zinc ion binding; IEA. | | | | |
| DR | GO; GO:0006730; F:one-carbon compound metabolism; IEA. | | | | |
| DR | InterPro; IPR01148; Euk_CoAhd. | | | | |
| DR | PFam; PF00194; Carb_anhydride; 1. | | | | |
| DR | ProDom; PD000865; Euk_CoAhd; 1. | | | | |
| DR | PROSITE; PS00162; EUK_CO2_ANHYDRASE; 1. | | | | |
| SQ | SEQUENCE 208 AA; 23451 MW; AF33D016A CO2_ANHYDRASE; 1. | | | | |
| SEQUENCE | 208 AA; 23451 MW; AF33D016A CO2_ANHYDRASE; 1. | | | | |
| Query Match | 42.7%; Score 551; DB 4; Length 208; | | | | |
| Best Local Similarity | 56.2%; Pred. No. 3.5e-41; | | | | |
| Matches | 100; Conservative 34; Mismatches 44; Indels 0; Gaps 0; | | | | |
| QY | 60 ISNSGHSPNVDFDDTENKSVLRGCPPLTSYRQLQVHLWGSADDHGSEHIVDGVSYAAELHVTAN 119 | | | | |
| Db | 5 ITNNGHSVQDVFDNFSDDRTVVGPLEGPYRILQFHFWGKHDVGSEHTVDGKSFPSBL 64 | | | | |
| QY | 120 HVVHWNDSKYPSTYEAAHEPDGLFLAVLGVLQFQGPNSQLOKHTDLSIKEKGKQTPTN 179 | | | | |
| Db | 65 HLVTHWNACKYSTGEAAASAPDGIAVAVGVFVLETSDEHPSPMNRLTDAVLYVRFGTKAQPSC 124 | | | | |
| QY | 180 FDLSLILPPSWDWTYPPSLSLTVPDILLESVTWIVLKOPINISSQOLAKERSLCTAEGE 237 | | | | |
| Db | 125 FNPKCLLPASRHYWTYPPSLSLTTPLESVTWIVLREPICISERQMKGKFRSLILPTSEDD 182 | | | | |
| RESULTS 10 | | | | | |
| Q9XZG6 | PRELIMINARY; PRT; 261 AA. | | | | |
| ID | Q9XZG6 | | | | |
| AC | Q9XZG6; | | | | |
| DT | 01-NOV-1999 (TREMBLrel. 12, Created) | | | | |
| DT | 01-NOV-1999 (TREMBLrel. 12, Last sequence update) | | | | |
| DR | 01-OCT-2003 (TREMBLrel. 25, Last annotation update) | | | | |
| DE | Carboxylic anhydride (BC 4.2.1.1); (Carbonate dehydratase). | | | | |
| OS | Anthopleura elegantissima (Sea anemone). | | | | |
| NCBI_TaxID | 6110; | | | | |
| RN | SEQUENCE FROM N.A. | | | | |
| RX | Weis V.M., Reynolds W.S.; "Carbonic anhydrase expression and synthesis in the sea anemone Anthopleura elegantissima are enhanced by the presence of diimide-like symbionts." Physiol. Biochem. Zool. 72:307-316 (1999). | | | | |
| RP | SEQUENCE FROM N.A. MEDLINE=19340407; PubMed=10222325; | | | | |
| RA | Weis V.M., Reynolds W.S.; "Carbonic anhydrase expression and synthesis in the sea anemone Anthopleura elegantissima are enhanced by the presence of diimide-like symbionts." Physiol. Biochem. Zool. 72:307-316 (1999). | | | | |
| RT | Antiochaea; Metazoa; Cnidaria; Anthozoa; Zoantharia; Anthopleura. | | | | |
| RT | Actiniidae; Anthopleura eleganssima (Sea anemone). | | | | |
| CC | NCBI_TaxID=6110; | | | | |
| CC | SEQUENCE FROM N.A. MEDLINE=19340407; PubMed=10222325; | | | | |
| CC | "Carbonic anhydrase expression and synthesis in the sea anemone Anthopleura elegantissima are enhanced by the presence of diimide-like symbionts." Physiol. Biochem. Zool. 72:307-316 (1999). | | | | |
| CC | - FUNCTION: REVERSIBLE HYDRATION OF CARBON DIOXIDE (By SIMILARITY). | | | | |
| CC | CATALYTIC ACTIVITY: H(2)CO (3) = CO (2) + H(2)O. | | | | |
| CC | SIMILARITY: BELONGS TO THE EUKARYOTIC-TYPE CARBONIC ANHYDRASE FAMILY. | | | | |

| | | | |
|-----------|---|--|---|
| DR | EMBL; AF140537; AAD32675.1; | - | |
| DR | HSSP; P0098; 1BV3. | - | |
| DR | GO; GO:0004089; F:carbonate dehydratase activity; IEA. | - | |
| DR | GO; GO:0016829; F:lyase activity; IEA. | - | |
| DR | GO; GO:0008270; F:zinc ion binding; IEA. | - | |
| DR | GO; GO:0006730; P:one-carbon compound metabolism; IEA. | - | |
| DR | InterPro; IPR001148; Buk_CoAhd. | - | |
| DR | Pfam; PF00194; carb_anhydride; 1. | - | |
| DR | ProDom; PD000865; Euk_COAhd; 1. | - | |
| DR | PROSITE; PS00162; Euk_CO2_anhydase; 1. | - | |
| KW | Lysine; Zinc. | - | |
| SQ | SEQUENCE: 261 AA; | 28612 MW; | 9CE4C82065DD9740 CRC64; |
| QY | 1 MSRLSGCYREHNGPIHWKEFFIPIADGDOQSPLIEKTKVYDSSL--RPLSIKYDSSAK 58 | Score 548.5; DB 5; Length 261; | |
| Db | 1 MAAPKWCYGPNNGPSKWAQFPAAGARQSPLDIKTHDAQHDSALKPKLQIQSQGNDF 60 | Best Local Similarity 46.6%; Fred. No. 8.1e-41; Matches 109; Conservative 43; Mismatches 77; Indels 5; Gaps 4; | SEQUENCE FROM N.A. |
| Qy | 59 IISNSGHSHFYNDDETDENKSVRIGGLPTGSTRLRQVHLWSADDHGSEHTVDGVSYAAE 118 | Score 548.5; DB 5; Length 261; | STRAIN=Berkeley; PMID=10471707; |
| Db | 61 NYTNNGKYSLVNKSRTKSEGTV-LSGGPLEHNRYEQFFHGWKTSGGGSEHLLDGKFAPE 119 | Best Local Similarity 46.6%; Fred. No. 8.1e-41; Matches 109; Conservative 43; Mismatches 77; Indels 5; Gaps 4; | RA |
| Qy | 119 LHVVHWNDSKYESFVEAHEPDKGLAVGFQLOQGEPSNQLQKITDTLSSTEKG-KOTRF 177 | Score 548.5; DB 5; Length 261; | Ashburner M., Misra S., Foote J., Lewis S.E., Blazej R., Davis T., Doyle C.M., Hwang Z.-Y., Wasserman D.A., Weinstein B., Wang A.H., Weissenbach J., Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A., Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L., Zheng X.H., Zhong F.W., Zhou X., Zhu S., Zhu X., Smith H.O., Svirskas R., Teeter C., Turner R., Venter J.C.; "The genome sequence of Drosophila melanogaster."; Science 287:2183-2195 (2000). [2] |
| Db | 120 LHLVHWTDLFSEFGEEASNSKNGLAGVAFQIGGESAAGLKHITDILPQVNIGDQDLKK 179 | Score 548.5; DB 5; Length 261; | RA |
| Qy | 178 TNFDLISLPPSW-DWYTPESSLTVPPLLESTWIVLKOPINISSQOLAKPRL 230 | Score 548.5; DB 5; Length 261; | RA |
| Db | 180 VPENLSSLLPENTNDTWYSGSLTTPCYEVSWFVKEPHTHENQMOQRSL 233 | Score 548.5; DB 5; Length 261; | RA |
| QY | 59 IISNSGHSHFYNDDETDENKSVRIGGLPTGSTRLRQVHLWSADDHGSEHTVDGVSYAAE 118 | Score 548.5; DB 5; Length 261; | SEQUENCE FROM N.A. |
| Db | 61 NYTNNGKYSLVNKSRTKSEGTV-LSGGPLEHNRYEQFFHGWKTSGGGSEHLLDGKFAPE 119 | Score 548.5; DB 5; Length 261; | STRAIN=Berkeley; |
| Qy | 119 LHVVHWNDSKYESFVEAHEPDKGLAVGFQLOQGEPSNQLQKITDTLSSTEKG-KOTRF 177 | Score 548.5; DB 5; Length 261; | RA |
| Db | 120 LHLVHWTDLFSEFGEEASNSKNGLAGVAFQIGGESAAGLKHITDILPQVNIGDQDLKK 179 | Score 548.5; DB 5; Length 261; | RA |
| Qy | 178 TNFDLISLPPSW-DWYTPESSLTVPPLLESTWIVLKOPINISSQOLAKPRL 230 | Score 548.5; DB 5; Length 261; | RA |
| Db | 180 VPENLSSLLPENTNDTWYSGSLTTPCYEVSWFVKEPHTHENQMOQRSL 233 | Score 548.5; DB 5; Length 261; | RA |
| RESULT 11 | | | SEQUENCE FROM N.A. |
| Q9V396 | PRELIMINARY; | PRT; 270 AA. | STRAIN=Berkeley; |
| LD | Q9V396 | PRT; 270 AA. | RA |
| AC | 01-MAY-2000 (T-EMBL; 13, Created) | RA | |
| DT | 01-MAY-2000 (T-EMBL; 13, Last sequence update) | RA | |
| DT | 01-OCT-2003 (T-EMBL; 25, Last annotation update) | RA | |
| QY | CAH1 protein (EC 4.2.1.1) (Carbonic anhydrase) (Carbonate dehydratase). | RA | |
| DE | DEHYDRASE | RA | |
| GN | CG7820.1 OR BD00941.1 | RA | |
| OS | Drosophila melanogaster (Fruit fly) | RA | |
| OC | Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydrioidae; Drosophilidae; Drosophila. | RA | |
| OC | NCBI_TaxID:7227; | RA | |
| RN | [1] | RA | |
| RP | SEQUENCE FROM N.A. | RA | |
| RC | STRAIN=Berkeley; | RA | |
| RC | MEDLINE=20196006; PubMed=10731132; | RA | |
| RA | Adams M.D., Celinker S.E., Holt R.A., Evans C.A., Gocayne J.D., Amanatides P.G., Scheerer S.E., Li P.W., Hoskins R.A., Galle R.F., Sutcliffe G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X., Brandon R.C., Rogers Y.H.C., Blazej R.G., Champé M.R., Pfeiffer B.D., Wan K.H., Apblett C., Baxter E.G., Dahlke C.R., Miklos G.I.J., Abril J.F., Aoyabe A., An H.-J., Richards S., Ashburner M., Henderson S.N., Baliew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M., Beeson K.Y., Benos P.V., Berman B.P., Bhanda D., Bolshakov S., Borckya D., Borchan M.R., Bouck J., Brokstein P., Brottier P., Burts K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I., Cherry J.M., Cawley S., Dahle C., Davenport L.B., Davies P., de Pablo B., Delcher A., Deng Z., Mayes A.D., Dev I., Dietz S.M., Dodson K.J., Douc L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P., Durbin R.K.J., Evans G.C., Ferreria S., Fleschmann W., Fosler C., Gabrilian A.E., Garg N.S., Gelbart W.M., Glasser K., Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M., Harris N.L., Harvey D., Heimann T.J., Hernandez J.R., Houck J., Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibeagwam C., Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A., | SEQUENCE 270 AA; 29952 MW; FEB85FC43800354 CRC64; Score 547.; DB 5; Length 270; Query Match 42.4%; Score 547.; DB 5; Length 270; | |

| | | | |
|--|---|--------|--|
| Best Local Similarity 49.3%; Pred. No. 1.2e-40; | Matches 113; Conservative 24; Mismatches 86; Indels 6; Gaps 4; | RN | SEQUENCE FROM N.A. |
| Matches 113; | Conservative 24; | RP | SEQUENCE FROM N.A. TISSUE=Kidney; |
| 6 WGYREHNGPIHKKEFFPIAEDQDQSPLIEKTKVEKYDSLR-PLSIKYDPESSAKLINS 63 | STRAIN=C57BL/6J; PubMed=11042159; | RC | STRAIN=C57BL/6J; TISSUE=Kidney; |
| 5 WGYTEENGPAWAKEYEQASGERQSPVDITPSSAKKGSEUNVAPLWKYYEHTKSLVNP 64 | MEDLINE=20499374; PubMed=11078661; | RX | MEDLINE=20499374; PubMed=11078661; |
| Db | Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M., Kono H., Okazaki Y., Muramatsu M., Hayashizaki Y.; "Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes." Genom Res. 10:11617-11630(2000). | RA | Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M., Kono H., Okazaki Y., Muramatsu M., Hayashizaki Y.; "Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes." Genom Res. 10:11617-11630(2000). |
| Qy | [5] | RN | [6] |
| 64 GHFSNVDDDEIENKSVLGGPLTGS-YRLRQVHLHGSADDHGSHHIVDGSYSYAAELHVY 122 | SEQUENCE FROM N.A. | RP | SEQUENCE FROM N.A. |
| 65 GYCWRDGVNDGVA-SEETGGPLGDQFLIKEFQFHCHWGCTDSKGSEHTVDGTSYSGFLHLV 122 | SEQUENCE FROM N.A. | RC | SEQUENCE FROM N.A. |
| Db | 123 HWNSDKYPFSVVAHEPDGLAVLGFLFQIGEBNSQLOKITDLSIKERG-KQTTRPTNFD 181 | RX | MEDLINE=2030913; PubMed=11078661; |
| Qy | 123 HWNNTKIKSFGEAAADPGLAVLGFLKAGNHAEILDKVTSILQFVILHGDRVTLPQGCD 182 | RA | Shibata K., Itoh M., Aizawa K., Kitisunai T., Tashiro H., Itoh M., Sumi N., Ishii Y., Nakamura S., Harada A., Nishine T., Harada A., Kono H., Akiyama J., Niizumi K., Kitisunai T., Tashiro H., Itoh M., Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K., Fujikawa S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M., Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsuzawa S., Kawai J., Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.; "RIKEN integrated sequence analysis (RISA) system: 3.84-format sequencing pipeline with 384 multicapillary sequencer." Genome Res. 10:1757-1771 (2000). |
| Db | 182 LLSLLPPSWDXWTYPGSITVPPLLESYTWIVLKOPINISSOQLAKFRSL 230 | RA | DR |
| Qy | 183 PGQLLPDPVHTWYTVEGSSLTPPCESESYTWIVFKTPLEVSDDQJNAMRNLL 231 | RA | MDG; MGI:88239; Ctr2. |
| Db | CAR2. | RA | GO; GO:0004820; F:carbonate dehydratase activity; IEA. |
| Q9DCT3 | PRELIMINARY; PRT; 192 AA. | RA | GO; GO:0006730; P:one-carbon compound metabolism; IEA. |
| ID | Q9DCT3; AC | RA | InterPro; IPR001148; Bk: Coanhds. |
| DT | 01-JUN-2001 (TREMBLrel. 17; Created) | RA | DR |
| DT | 01-MAR-2003 (TREMBLrel. 23; Last sequence update) | RA | Pfam; PF00194; carb_anhydrase_1. |
| DT | 01-OCT-2003 (TREMBLrel. 25; Last annotation update) | RA | ProdDom; PD000865; Bjk_Coanhds_1. |
| DE | Adult male kidney cDNA. RIKEN full-length enriched library, clone:061010L11 product:carbonic anhydrase 2, full insert sequence | RA | DR |
| DE | (Fragment). | RA | PROSITE; PS00162; BUK_CO2_ANHYDRASE; 1. |
| GN | | FT | NON_TER 1 1 |
| OS | Mus musculus (Mouse). | SQ | SEQUENCE 192 AA; 21486 MW; E7E01D0E98692043 CRC64; |
| OC | Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi; | RN | Query Match 42.4%; Score 546.5; DB 11; Length 192; |
| OC | Mammalia; Eutheria; Rodentia; Sciurognath; Muridae; Murinae; Mus. | RN | Best Local Similarity 59.8%; Pred. No. 7.9e-41; |
| OX | | RN | Matches 101; Conservative 28; Mismatches 39; Indels 1; Gaps 1; |
| RN | SEQUENCE FROM N.A. | QY | 70 DEDDTENKSVLRGGPLTSYRQLRQVHLHWGSADDHGSEHIVDGSYSYAAELHVY 129 |
| [1] | NCBI_TAXID=1090; | DB | 1 EFDDSDQNAVLRKGPSPSRSYRQLQHFHWGSSQGSHTVNKRYAAELHVWNT-KY 59 |
| RP | SEQUENCE FROM N.A. | QY | 130 PSFVEAAHPDGIAVGLGYFLQCEPNSQLQKTTDLSIKEKGKQTRFTNFDLSSLLPPS 189 |
| RC | STRAIN=C57BL/6J; TISSUE=Kidney; | DB | 60 GDFGKAVQPDGLIAVGLFLKIGPASQQLQKVLEALISIKTKRKAFAANFDPSLLPGN 119 |
| RA | Adachi J., Aizawa K., Akahira S., Akimura T., Arai A., Ando H., Arakawa T., Bono H., Carninci P., Fukuda S., Furuno M., Hanagaki T., Hara A., Hayatani N., Hiramoto K., Hiraoka T., Horii F., Imotani K., Ishii Y., Itoh M., Izawa M., Kasukawa T., Kato H., Matsuyama T., Miayazaki A., Nishi K., Nomura K., Nunazaki R., Ohno M., Okido T., Owa H., Saito H., Saito R., Sakai C., Sakai K., Sano H., Sasaki D., Shiba K., Shitaya A., Shiraki T., Sogabe Y., Suzuki H., Tagami M., Tagawa A., Takahashi F., Tanaka T., Tejima Y., Toya T., Yanamura T., Yasunishi A., Yoshida K., Yoshino M., Muramatsu M., Hayashizaki Y.; Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases. | RA | 190 WDWTYPSGLTVPPPLESTWVLPQINISSQQLAKERSLLCTAGEA 238 |
| RN | SEQUENCE FROM N.A. | QY | 120 LDWTYPSLTTPLLCYTWIREPITVSSEBOMSHRTLNNEEGDA 168 |
| RN | SEQUENCE FROM N.A. | DB | |
| RC | SEQUENCE FROM N.A. TISSUE=Kidney; MEDLINE=22354683; PubMed=112466851; | QBMFBH | RESULT 1.3 |
| RX | The FANTOM Consortium, the RIKEN Genome Exploration Research Group Phase I & II Team; "Analysis of the mouse transcriptome based on functional annotation of 60-770 full-length cDNAs." ; Nature 420:563-573 (2002). | ID | OBMPHB PRELIMINARY; PRT; 243 AA. |
| RA | RN | AC | QBMPHB; QBMFBH; DT 01-OCT-2002 (TREMBLrel. 22, Created) |
| RA | "Functional annotation of a full-length mouse cDNA collection." ; Nature 409: 685-690 (2000). | DT | DT 01-OCT-2002 (TREMBLrel. 22, Last sequence update) |
| RA | RN | DE | DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update) |
| RA | RN | GN | DE Catheionic anhydrase (EC 4.2.1.1). |
| RA | RN | CA | GN CAL |
| RA | RN | OS | Riftia pachyptila (Tube worm). |
| RA | RN | OC | Eukaryota; Metazoa; Pooionphora; Vestimentifera; Anellobanchia; Riftiidae; Riftiidae; Riftia. |
| RA | RN | OX | NCBI_TaxID=6426; |
| RA | RN | RN | SEQUENCE FROM N.A. |
| RA | RN | RA | De Cian M.C., Baillie X., Boulben S., Strub J.M., Von Dorssalaer A., Laillier P.H.; "An insight into molecular and biochemical characteristics of carbonic anhydrases from Riftia pachyptila, a symbiotic invertebrate living under extreme conditions"; "High-efficiency full-length cDNA cloning." ; Meth. Enzymol. 303:19-44 (1999). |
| RA | RN | RA | Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases. |

| | | | | |
|--|---|--|--|---|
| DR | EMBL; AJ439711; CAD29128.1; | | Qy | 122 VHNNSDKYPSPFVEAAHFPDGLAVLGVLQIGEPNSSLQQLQKITYDKEKGKQTRFTNFD 181 |
| DR | GO; GO:001829; F:carbonate dehydratase activity; IEA. | | Db | 148 VHYNSDLYPDFESTASDKSEGLAVLIEIGSANPSYDKIFSHLQAVKYKGQVLIQGFN 207 |
| DR | GO; GO:0008270; F:zinc ion binding; IEA. | | | |
| DR | GO; GO:0006730; P:one-carbon compound metabolism; IEA. | | | |
| DR | InterPro; IPR001148; Euk_Coanh. | | Qy | 182 LISSLPPS-WDYWTYPSLTVPLLESVTWIVLKOPINISSCOLAKFRSLI 231 |
| DR | InterPro; IPR001149; carb_anhydrole. | | | |
| DR | Pfam; PF00194; carb_anhydrole. | | | |
| DR | ProDom; PD000865; Euk_Coanh. | | | |
| DR | PROSITE; PS00162; EUR_CO2_ANHYDRASE_1. | | | |
| KW | KW | | | |
| SEQUENCE | 243 AA; 25661 MW; AAA53B407A1F4645 CRC64; | | SQ | 208 IEELPPSPGEYRYEBSLTPPCPVLYTFRNPVQISOBLLAETAL 258 |
| <hr/> | | | | |
| RESULT 15 | | | | |
| Qy | 093587 PRELIMINARY; ID: 093587; AC: 093587; | | Qy | 093587 PRELIMINARY; ID: 093587; AC: 093587; |
| Best Local Similarity 37.8%; Score 487.5; Length 243; | Pred. No. 2.1e-35; Gaps 5; | | Best Local Similarity 37.8%; Score 487.5; Length 243; | Pred. No. 2.1e-35; Gaps 5; |
| Matches 103; Conservative 25; Mismatches 91; Indels 9; Gaps 5; | | | Matches 103; Conservative 25; Mismatches 91; Indels 9; Gaps 5; | |
| Db | 5 SWGYREHNGPIHWKEFFPFIADGQDQSPLEIKTKVEVKYDSLRPLSIKYDPSAKILNSG 64 | | Db | 5 SWGYREHNGPIHWKEFFPFIADGQDQSPLEIKTKVEVKYDSLRPLSIKYDPSAKILNSG 64 |
| Qy | 3 AWDY-EANGPATWAKSFPLAAKGKQSPIDPAVS -KKSTSALVASYNPARASNTUTNTG 60 | | Qy | 3 AWDY-EANGPATWAKSFPLAAKGKQSPIDPAVS -KKSTSALVASYNPARASNTUTNTG 60 |
| Db | 65 HSFNVDEDDTENKSVLPGGPLTSYRQVHWHGSADDHGSEHIVGVSAAELHVH 124 | | Db | 65 HSFNVDEDDTENKSVLPGGPLTSYRQVHWHGSADDHGSEHIVGVSAAELHVH 124 |
| Qy | 61 LSFQVSTDTGTC----LSGGPGLNEYKAASFEHWWSKTSAGESEHTVAGKAYAAEHVH 115 | | Qy | 61 LSFQVSTDTGTC----LSGGPGLNEYKAASFEHWWSKTSAGESEHTVAGKAYAAEHVH 115 |
| Db | 125 NSDKYPSEVAHEPDGLAVLGVLQFQGPNSQQLKITYDKEKGKQTRFT-NFDL 183 | | Db | 125 NSDKYPSEVAHEPDGLAVLGVLQFQGPNSQQLKITYDKEKGKQTRFT-NFDL 183 |
| Qy | 116 NAAKYASQDADGGLAVLAATEFIQGANTHVGQKTLIDLPSPVKGDTATIPGFDVA 175 | | Db | 116 NAAKYASQDADGGLAVLAATEFIQGANTHVGQKTLIDLPSPVKGDTATIPGFDVA 175 |
| Db | 184 SLIPPSWDM-YWTYPSLTVPLLESVTWIVLKOPINISSCOLAKFRSL 230 | | Qy | 184 SLIPPSWDM-YWTYPSLTVPLLESVTWIVLKOPINISSCOLAKFRSL 230 |
| Qy | 176 CLIPGDOSKQWYYPGSLITPPCFESVTWIVKDPQLCENQLAALRKI 223 | | Db | 176 CLIPGDOSKQWYYPGSLITPPCFESVTWIVKDPQLCENQLAALRKI 223 |
| <hr/> | | | | |
| RESULT 14 | | | | |
| Q8K2J1 | PRELIMINARY; ID: Q8K2J1; AC: Q8K2J1; Version: 22, Created: 01-OCT-2002 (TRMBLrel. 22, Last sequence update) | | Qy | 4 LSWGYREHNGPIHWKEFFPFIADGQDQSPLEIKTKVEVKYDSLRPLSIKYDPSAKILNS 63 |
| DR | DT 01-OCT-2002 (TRMBLrel. 22, Last sequence update) | | Db | 1 MSWGYADNGPDKWAENPYANGPROSPIDLPGDAEDALKPLSKYDPTARTILNN 60 |
| DR | DT 01-OCT-2003 (TRMBLrel. 25, Last annotation update) | | Qy | 64 GHSENNDFDDTENKSVLPGGPLTSYRQVHWHGSADDHGSEHIVGVSAAELHVH 123 |
| DE | Hypothetical protein. | | Db | 61 GHSPFQTTFIDDTDSSTLKDGTISGVYRLQKFHHMGACDEKGSEVTAG----- 109 |
| OS | Mus musculus (Mouse). | | Qy | 65 WNSDKYPSPFVEAAHEDGLAVLGVLQIGEPBNSQ-----LQKIDTP-LDSIKEK-GKQTR 176 |
| OC | Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus. | | Db | 110 -NMCIIPISSISCT-----GTPNTRAISEKPLANLTDSPLSEFSSKGULKTP 153 |
| NCBI_TAXID-10090; | OX OX | | Qy | 177 FTNFEDLISLPLPPSWDWTYPSL-TVPPILLESVTW-----IVLKQ 215 |
| RN | [1] | | Db | 154 -TSRRRLTPSPSRPASPPSLATPPACSPGAWTTGSKKAPGPIPLCWRSPIGILSARE 212 |
| RP | SEQUENCE FROM N.A. | | Qy | 216 PINISSOQLAKFRSLICTAREA 238 |
| RA | Strausberg R; | | Db | 213 PISVSAEQMAKFRSLIFSAGEA 235 |
| RL | Submitted (JUN-2002) to the EMBL/GenBank/DBJ databases. | | | |
| DR | EMBL; BC031385; AAH31385.1; MGD; MGI:1923709; 2310047E01Rik. | | | |
| DR | GO; GO:0008270; F:carbonate dehydratase activity; IEA. | | | |
| DR | GO; GO:0006730; P:one-carbon compound metabolism; IEA. | | | |
| DR | InterPro; IPR001148; Euk_Coanh. | | | |
| DR | Pfam; PF00194; carb_anhydrole. | | | |
| DR | ProDom; PD000865; Euk_Coanh. | | | |
| DR | PROSITE; PS00162; EUR_CO2_ANHYDRASE_1. | | | |
| KW | Hypothetical protein. | | | |
| SEQUENCE | 344 AA; 38724 MW; 12DCBD6C3B45D75A CRC64; | | | |
| <hr/> | | | | |
| Query Match 33.9%; Score 436.5; DB 11; Length 344; | Best Local Similarity 40.3%; Pred. No. 1.2e-30; Mismatches 90; Indels 9; Gaps 5; | | Search completed: September 9, 2004, 15:13:55 | |
| Matches 93; Conservative 39; Mismatches 90; Indels 9; Gaps 5; | Job time : 118 secs | | | |
| Qy | 6 WGYREHNGPIHWKEFFPFIADGQDQSPLEIKTKVEVKYDSLRPLSIKYDPSAKI--ISN 62 | | | |
| Db | 32 WTYVGPAGEKONWSKYPSCGGPLQDPLHDSAPLQFGVYNSVEKLNLN 91 | | | |
| Qy | 63 SGHSFNVDDTENKSVLPGGPLTSYRQVHWHGSADD-HGSEHIVGVSAAELHV 121 | | | |
| Db | 92 DGHSVRNLN---NSDMYIQGLQPHHYRAEQLHLHMGNRNDPHGSEHTVSGKFAELHV 147 | | | |

| Result No. | Score | Query Match Length | DB ID | Description |
|------------|-------|--------------------|-------|--|
| 1 | 804 | 62.4 | 261 | 1 US-08-015-986A-8 Sequence 8, Appli |
| 2 | 804 | 62.4 | 261 | 2 US-08-0446-363-12 Sequence 8, Appli |
| 3 | 802 | 62.2 | 260 | 1 US-08-015-973-4 Sequence 4, Appli |
| 4 | 802 | 62.2 | 260 | 2 US-08-081-164-4 Sequence 4, Appli |
| 5 | 802 | 62.2 | 260 | 3 US-08-081-929-4 Sequence 4, Appli |
| 6 | 788.5 | 61.2 | 259 | 1 US-08-015-973-5 Sequence 5, Appli |
| 7 | 788.5 | 61.2 | 259 | 2 US-08-044-164-6 Sequence 5, Appli |
| 8 | 788.5 | 61.2 | 259 | 3 US-08-081-929-5 Sequence 5, Appli |
| 9 | 788.5 | 61.2 | 260 | 1 US-08-015-986A-9 Sequence 9, Appli |
| 10 | 788.5 | 61.2 | 260 | 3 US-08-044-363-9 Sequence 9, Appli |
| 11 | 788.5 | 61.2 | 260 | 3 US-08-027-128-4 Sequence 4, Appli |
| 12 | 788.5 | 61.2 | 260 | 4 US-09-976-59-644 Sequence 644, App |
| 13 | 788.5 | 61.2 | 288 | 2 US-08-095-886C-12 Sequence 6, Appli |
| 14 | 788.5 | 61.2 | 288 | 3 US-09-039-819A-12 Sequence 12, Appli |
| 15 | 788.5 | 61.2 | 288 | 4 US-09-750-911-12 Sequence 12, Appli |
| 16 | 748.5 | 58.1 | 259 | 1 US-08-015-986A-10 Sequence 10, Appli |
| 17 | 748.5 | 58.1 | 259 | 1 US-08-015-973-6 Sequence 6, Appli |
| 18 | 748.5 | 58.1 | 259 | 2 US-08-044-363-10 Sequence 10, Appli |
| 19 | 748.5 | 58.1 | 259 | 2 US-08-044-164-6 Sequence 6, Appli |
| 20 | 748.5 | 58.1 | 259 | 3 US-08-081-929-6 Sequence 1, Appli |
| 21 | 747.5 | 58.0 | 421 | 4 US-09-938-270B-1 Sequence 18, Appli |
| 22 | 695 | 53.9 | 201 | 1 US-08-016-667-18 Sequence 18, Appli |
| 23 | 695 | 53.9 | 201 | 1 US-09-024-110-18 Sequence 18, Appli |
| 24 | 695 | 53.9 | 201 | 5 PCT-US5-0789-18 Sequence 18, Appli |
| 25 | 668 | 51.8 | 261 | 1 US-08-015-973-9 Sequence 9, Appli |
| 26 | 668 | 51.8 | 261 | 2 US-08-044-164-9 Sequence 9, Appli |
| 27 | 668 | 51.8 | 261 | 3 US-08-081-929-9 Sequence 9, Appli |

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.
OM protein - protein search, using sw model
Run on: September 9, 2004, 15:09:25 ; Search time 32 Seconds
(without alignments)
390.421 Million cell updates/sec
Title: US-10-069-434-1
Perfect score: 1289
Sequence: 1 M S R L S W G Y R E H N G P I H K E F Q L A K F E R S U L C T A E G E A A F L 242
Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5
Searched: 389414 seqs, 51625971 residues
Total number of hits satisfying chosen parameters: 389414

ALIGNMENTS

RESULT 1
US-08-015-986A-8
; Sequence 8, Application US/08015986A
; Patent No. 5532123
; GENERAL INFORMATION:
; APPLICANT: Schlessinger, Joseph
; TITLE OF INVENTION: NOVEL RECEPTOR-TYPE PHOSPHOTYROSINE
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS: 14
; ADDRESSEE: BENNIE & EDMONDS
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: FLOPPY DISK
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/015,986A
; FILING DATE: 10-FEB-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Misrock, S. Leslie
; REGISTRATION NUMBER: 18,872
; REFERENCE/DOCKET NUMBER: 7683-028
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-790-9090
; TELEFAX: 212-869-8864/9741
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 261 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: unknown
; MOLECULE TYPE: Protein
; US-08-015-986A-8

SUMMARIES

| Query | Match Length | DB ID | Description |
|-------|--------------|---------------------|--------------------|
| 1 | 261 | 1 US-08-015-986A-8 | Sequence 8, Appli |
| 2 | 261 | 2 US-08-0446-363-8 | Sequence 8, Appli |
| 3 | 260 | 1 US-08-015-973-4 | Sequence 4, Appli |
| 4 | 260 | 2 US-08-081-164-4 | Sequence 4, Appli |
| 5 | 260 | 3 US-08-081-929-4 | Sequence 4, Appli |
| 6 | 259 | 1 US-08-015-973-5 | Sequence 5, Appli |
| 7 | 259 | 2 US-08-044-164-6 | Sequence 5, Appli |
| 8 | 259 | 3 US-08-081-929-5 | Sequence 5, Appli |
| 9 | 260 | 1 US-08-015-986A-9 | Sequence 9, Appli |
| 10 | 260 | 3 US-08-044-363-9 | Sequence 9, Appli |
| 11 | 260 | 3 US-08-027-128-4 | Sequence 4, Appli |
| 12 | 260 | 4 US-09-976-59-644 | Sequence 644, App |
| 13 | 288 | 2 US-08-095-886C-12 | Sequence 6, Appli |
| 14 | 288 | 3 US-09-039-819A-12 | Sequence 12, Appli |
| 15 | 288 | 4 US-09-750-911-12 | Sequence 12, Appli |
| 16 | 259 | 1 US-08-015-986A-10 | Sequence 10, Appli |
| 17 | 259 | 1 US-08-015-973-6 | Sequence 6, Appli |
| 18 | 259 | 2 US-08-044-363-10 | Sequence 10, Appli |
| 19 | 259 | 2 US-08-044-164-6 | Sequence 6, Appli |
| 20 | 259 | 3 US-08-081-929-6 | Sequence 1, Appli |
| 21 | 421 | 4 US-09-938-270B-1 | Sequence 18, Appli |
| 22 | 53.9 | 201 | 1 US-08-016-667-18 |
| 23 | 53.9 | 201 | 1 US-09-024-110-18 |
| 24 | 53.9 | 201 | 5 PCT-US5-0789-18 |
| 25 | 51.8 | 261 | 1 US-08-015-973-9 |
| 26 | 51.8 | 261 | 2 US-08-044-164-9 |
| 27 | 51.8 | 261 | 3 US-08-081-929-9 |

Qy 1 MSRLSGYREHNGPITHKEFFPIADGQQSPIEIKTKEYRDSSLRPLSIKYDPSAKII 60
Db 1 MASPDWGYDXKNGPQWSKUPIANGNNQSVDIKTSETXHDTSLKPISVSYNPATAKEI 60

QY 61 SNSGHSNVDDEDTENKSVLRGGLTGSYRQLQVHLHGSADDHGSEHIVDGSYAAELH 120
 Db 61 INVGHSFHVNFEDNDRSVLKGGPFDSYRLQFPHWGSTNEHGSSEHTDVGVKSAELH 120

QY 121 VVHNNSDKYPSVTEAAHEDGLAVLGYFLQIGEPNSQLOKITDLSIKEKGKOTRFTNF 180
 Db 121 VAHNSAKYSSLAASKADGLAVICVLMKVGEANPKQVKVLDALQAIKGKRAFTNF 180

RESULT 3
 US-08-015-973-4
 Sequence 4, Application US/08015973
 Patent No. 5614094
 GENERAL INFORMATION:
 APPLICANT: Schlessinger, Joseph
 TITLE OF INVENTION: A NOVEL RECEPTOR-TYPE PHOSPHOTYROSTEIN
 TITLE OF INVENTION: PHOSPHATASE-BETA
 NUMBER OF SEQUENCES: 9
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: PENNIE & EDMONDS
 STREET: 1155 Avenue of the Americas
 CITY: New York
 STATE: New York
 COUNTRY: U.S.A.
 ZIP: 10036-2711
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/446,363
 FILING DATE: 10-FEB-1993
 ATTORNEY/AGENT INFORMATION:
 NAME: Misrock, S. Leslie
 REGISTRATION NUMBER: 18,872
 REFERENCE/DOCKET NUMBER: 7683-021
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (212) 790-9020
 TELEFAX: (212) 869-8864/9741
 CLASSIFICATION: 435
 PRIORITY APPLICATION DATA:
 APPLICATION NUMBER: US 08/015,986
 FILING DATE: 10-FEB-1993
 ATTORNEY/AGENT INFORMATION:
 NAME: Misrock, S. Leslie
 REGISTRATION NUMBER: 18,872
 REFERENCE/DOCKET NUMBER: 7683-028
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 212-869-8864/9741
 TELEX: 66141 PENNIE
 INFORMATION FOR SEQ ID NO: 8:
 SEQENCE CHARACTERISTICS:
 LENGTH: 261 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: unknown
 MOLECULE TYPE: protein
 US-08-015-973-4

QY 181 DLLSLPPSPNDYWTYPPGSLTVPPPLIESVTWVLKOPINISQQQLAKERSLCTAEGEEA 239
 Db 181 DPSTLPPSLDFWTYPPGSLTHPPLESVTWILCKESISVSSEQIQAQRSLLSNVGDN 239

QY 181 DLLSLPPSPNDYWTYPPGSLTVPPPLIESVTWVLKOPINISQQQLAKERSLCTAEGEEA 239
 Db 181 DPSTLPPSLDFWTYPPGSLTHPPLESVTWILCKESISVSSEQIQAQRSLLSNVGDN 239

QY 186 LPPSDWYNTYPPGSLTVPPPLIESVTWVLKOPINISQQQLAKERSLCTAEGEEA 239
 Db 186 LPPSDWYNTYPPGSLTVPPPLIESVTWVLKOPINISQQQLAKERSLCTAEGEEA 239
 QY 185 LPSSLDFWTYPPGSLTHPPLESVTWILCKESISVSSEQIQAQRSLLSNVGDN 238
 Db 185 LPSSLDFWTYPPGSLTHPPLESVTWILCKESISVSSEQIQAQRSLLSNVGDN 238

RESULT 4
 US-08-448-164-4
 Sequence 4, Application US/08448164
 Patent No. 5925336

GENERAL INFORMATION:

APPLICANT: Schlesinger, Joseph
 TITLE OF INVENTION: A NOVEL RECEPTOR-TYPE PHOSPHOTYROSINE PHOSPHATASE-BETA

NUMBER OF SEQUENCES: 9

CORRESPONDENCE ADDRESS:

ADDRESSEE: PENNIE & EDMONDS
 STREET: 1155 Avenue of the Americas
 CITY: New York
 STATE: New York
 COUNTRY: U.S.A.
 ZIP: 10036

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent in Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/448,164
 FILING DATE: 24 MAY 1995
 CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08 / 015, 973
 FILING DATE: 10 FEB 1993

ATTORNEY/AGENT INFORMATION:

NAME: Misrock, S. Leslie
 REGISTRATION NUMBER: 18,972
 REFERENCE/DOCKET NUMBER: 7683-021

TELECOMMUNICATION INFORMATION:

TELEPHONE: (212) 869-3090
 TELEFAX: (212) 869-8644/9741
 TELEX: 66141 PENNIE
 INFORMATION FOR SEQ ID NO: 4:

SEQUENCE CHARACTERISTICS:
 LENGTH: 260 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: unknown
 MOLECULE TYPE: protein

US - 08 - 448 - 164 - 4

Query Match 62.2%; Score 802; DB 2; Length 260;
 Best Local Similarity 60.3%; Pred. No. 3.1e-76;
 Matches 141; Conservative 43; Mismatches 50; Indels 0; Gaps 0;

Qy 6 WGYREENGPIHMKKEPPIADGQSPTEITKTYKEVYKDSSLRPLPSIKEYDPSAKITSNSGH 65
 Db 5 WGYDDDRNGPEQWSKLPIANGNNQSPVDTIKSETKHTSLKPISSYNPATAKEIINVGH 64

Qy 66 SENVDDDDTENKSVLRGGLTGSYRLRQVHMGADDHGSEHIVDGVSAAELHVHN 125
 Db 65 SFHVNEDNDNRSVLKGPFSDSYRFQFHFWGSNEHGSBHTVGVRSAELVHN 124

Qy 126 SDKYPFVEAHEPDGLAVGFVFLQIGEPNSQLOKTTDLSIKEKGKOTRTNFDSL 185
 Db 125 SAKYSSLAEEASKADGLAVGVLKGEANPKLQKVDAQKTKGRKRAFTNFDPSTL 184

Qy 186 LPPSMWYWTFPGSLTPPLSITVWILKOPINISSQALKERSLICTAEGAA 239
 Db 185 LPSSLDFTWYTFPGSLTPPLYSVWICKESISVSEQLAQFRSLNSVNGDNA 238

RESULT 6
 US - 08 - 015 - 973 - 5

Qy 186 LPPSMWYWTFPGSLTPPLSITVWILKOPINISSQALKERSLICTAEGAA 239
 Db 185 LPSSLDFTWYTFPGSLTPPLYSVWICKESISVSEQLAQFRSLNSVNGDNA 238

RESULT 5
 US - 08 - 081 - 929 - 4

Sequence 4, Application US/08081929
 Patent No. 6160090

GENERAL INFORMATION:

APPLICANT: Schlesinger, Joseph
 APPLICANT: Barnea, Gilad
 APPLICANT: Grunet, Martin H.
 APPLICANT: Margolin, Richard U.

TITLE OF INVENTION: A NEW CLASS OF RPTPases: THEIR STRUCTURAL DOMAINS AND LIGANDS

COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent In Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/015, 973
 FILING DATE: 10-FEB-1993
 CLASSIFICATION: 435
 ATTORNEY/AGENT INFORMATION:
 NAME: Misrock, S. Leslie
 REGISTRATION NUMBER: 18,872
 REFERENCE/DOCKET NUMBER: 7683-021
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (212) 790-9090
 TELEX: (212) 869-8864/9741
 FAX: 66141 PENNIE
 TOPOLGY: unknown
 MOLECULE TYPE: protein
 US-08-015-973-5

SEQUENCE CHARACTERISTICS:
 LENGTH: 259 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 STRATEGY: unknown
 MOLECULE TYPE: protein
 US-08-048-164-5

Query Match 61.2%: Score 788.5; -DB 1; Length 259;
 Best Local Similarity 61.6%; Prod. No. 8e-75; Mismatches 33; Indels 1; Gaps 1;
 Matches 143; Conservative 33; Query 6 WGYREHNGPIHWKEFFPPIADGDOQSPIEIKTKEVYKDYSSRLPLSIKYDPPSSAKIISNGH 65
 Db 4 WGKGKHNGSEHHWKKDFPAKGERQSPVDDHTAKYDSLKPUSVSYOATSRLUNNGH 63

Query 66 SFNDFFDTONKSVLRGSPPLTSYRLEQVHLIWSADDHGSEHIVDGVSAYAELHVHN 125
 Db 64 AFNVEFDISQDAVKLKGGLDGTYRLQFHFWGSLDGQGSEHTVDCKKYAAELHVHN 123

Query 126 SDKPSFYEAHHEDGLAVLGFLQIGBNSQLQKITDTLDSTKEKGQTRTFNFDLISL 185
 Db 124 T-KYGDGEKAQODPGLAVLGFLKVGSAKPGLQKVWDVLDSTKTKGSADFTNFDPRLG 182

Query 66 LPPSWDWTYPGSLTVPLLESVTWILKOPINISSQQLAKERSLICTAEGE 237
 Db 64 AFNVEFDISQDAVKLKGGLDGTYRLQFHFWGSLDGQGSEHTVDCKCYAAELHVHN 123

Query 126 SDKPSFYEAHHEDGLAVLGFLQIGBNSQLQKITDTLDSTKEKGQTRTFNFDLISL 185
 Db 124 T-KYGDGEKAQODPGLAVLGFLKVGSAKPGLQKVWDVLDSTKTKGSADFTNFDPRLG 182

RESULT 8
 US-08-015-929-5
 Sequence 5, Application US/08081929
 Patent No. 6160090
 GENERAL INFORMATION:
 APPLICANT: Schlessinger, Joseph
 APPLICANT: Barnes, Gilad
 APPLICANT: Grumer, Martin H.
 APPLICANT: Margolis, Richard U.
 TITLE OF INVENTION: A NEW CLASS OF RPTPasses; THEIR
 TITLE OF INVENTION: STRUCTURAL DOMAINS AND LIGANDS
 NUMBER OF SEQUENCES: 13
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: PENNIE & EDMONDS
 STREET: 1155 Avenue of the Americas
 CITY: New York
 STATE: New York
 COUNTRY: U.S.A.
 ZIP: 10036

COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/448,164
 FILING DATE: 24-MAY-1995
 CLASSIFICATION: 435
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/015, 973
 FILING DATE: 10-FEB-1993

ATTORNEY/AGENT INFORMATION:
 NAME: Misrock, S. Leslie
 REGISTRATION NUMBER: 18,872
 REFERENCE/DOCKET NUMBER: 7683-021
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (212) 790-9090
 TELEX: (212) 869-8864/9741
 FAX: 66141 PENNIE
 INFORMATION FOR SEQ ID NO: 5:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 259 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 STRATEGY: unknown
 MOLECULE TYPE: protein
 US-08-048-164-5

Query Match 61.2%: Score 788.5; -DB 2; Length 259;
 Best Local Similarity 61.6%; Prod. No. 8e-75; Mismatches 33; Indels 1; Gaps 1;
 Matches 143; Conservative 33; Query 6 WGYREHNGPIHWKEFFPPIADGDOQSPIEIKTKEVYKDYSSRLPLSIKYDPPSSAKIISNGH 65
 Db 4 WGKGKHNGSEHHWKKDFPAKGERQSPVDDHTAKYDSLKPUSVSYOATSRLUNNGH 63

Query 66 SFNDFFDTONKSVLRGSPPLTSYRLEQVHLIWSADDHGSEHIVDGVSAYAELHVHN 125
 Db 64 AFNVEFDISQDAVKLKGGLDGTYRLQFHFWGSLDGQGSEHTVDCKKYAAELHVHN 123

Query 126 SDKPSFYEAHHEDGLAVLGFLQIGBNSQLQKITDTLDSTKEKGQTRTFNFDLISL 185
 Db 124 T-KYGDGEKAQODPGLAVLGFLKVGSAKPGLQKVWDVLDSTKTKGSADFTNFDPRLG 182

Query 66 LPPSWDWTYPGSLTVPLLESVTWILKOPINISSQQLAKERSLICTAEGE 237
 Db 64 AFNVEFDISQDAVKLKGGLDGTYRLQFHFWGSLDGQGSEHTVDCKCYAAELHVHN 123

Query 126 SDKPSFYEAHHEDGLAVLGFLQIGBNSQLQKITDTLDSTKEKGQTRTFNFDLISL 185
 Db 124 T-KYGDGEKAQODPGLAVLGFLKVGSAKPGLQKVWDVLDSTKTKGSADFTNFDPRLG 182

RESULT 8
 US-08-015-929-5
 Sequence 5, Application US/08081929
 Patent No. 6160090
 GENERAL INFORMATION:
 APPLICANT: Schlessinger, Joseph
 APPLICANT: Barnes, Gilad
 APPLICANT: Grumer, Martin H.
 APPLICANT: Margolis, Richard U.
 TITLE OF INVENTION: A NEW CLASS OF RPTPasses; THEIR
 TITLE OF INVENTION: STRUCTURAL DOMAINS AND LIGANDS
 NUMBER OF SEQUENCES: 13
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: PENNIE & EDMONDS
 STREET: 1155 Avenue of the Americas
 CITY: New York
 STATE: New York
 COUNTRY: U.S.A.
 ZIP: 10036

COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/081,929
 FILING DATE: 23-JUN-1993
 CLASSIFICATION: 514
 ATTORNEY/AGENT INFORMATION:
 NAME: Corzzi, Laura A.
 REGISTRATION NUMBER: 30742
 TELECOMMUNICATION INFORMATION:
 REFERENCE/DOCKET NUMBER: 7683-041-999
 TELEPHONE: 212 790-9090
 TELEX: 212 869-8864/9741

TELEX: 66141 PENNIE
 / INFORMATION FOR SEQ ID NO: 5:
 / SEQUENCE CHARACTERISTICS:
 / LENGTH: 259 amino acids
 / TYPE: amino acid
 / STRANDEDNESS: single
 / TOPOLOGY: unknown
 / MOLECULE TYPE: protein
 US-08-081-929-5

Query Match Best Local Similarity 61.2%; Score 788.5; DB 1; Length 260;
 Matches 143; Conservative 33; Mismatches 55; Indels 1; Gaps 1;

```

  Qy   6 WGYREHNGPIHWKEFFPPIADGDOQSPIEIKTKEYVKKDSSURPLSIKYDPSSAKIISNSGH 65
  Db   5 WGYCKHNGPEHWKHDFTAKGERQSPVLDTHTAKYDPSLKPVLSVYDQATSLRILNNGH 64
  
```

Query Match Best Local Similarity 61.6%; Score 788.5; DB 1; Length 259;
 Matches 143; Conservative 33; Mismatches 55; Indels 1; Gaps 1;

```

  Qy   6 WGYREHNGPIHWKEFFPPIADGDOQSPIEIKTKEYVKKDSSURPLSIKYDPSSAKIISNSGH 65
  Db   4 WGYCKHNGPEHWKHDFTAKGERQSPVLDTHTAKYDPSLKPVLSVYDQATSLRILNNGH 63
  
```

Query Match Best Local Similarity 61.6%; Score 788.5; DB 1; Length 259;
 Matches 143; Conservative 33; Mismatches 55; Indels 1; Gaps 1;

```

  Qy   6 SENVDDDETEENKVSLRGGLTGSYRLRQVHLWMSADDHGSEHTVDGVSYAELHVHN 125
  Db   6 SENVDDDETEENKVSLRGGLTGSYRLRQVHLWMSADDHGSEHTVDGVSYAELHVHN 125
  
```

Query Match Best Local Similarity 61.6%; Score 788.5; DB 1; Length 259;
 Matches 143; Conservative 33; Mismatches 55; Indels 1; Gaps 1;

```

  Qy   6 AFNVEEDDSQDAKVLKGGLDGTYRLQFHFHWGSLDGQSEHTVDKKYAAELHVHN 124
  Db   6 AFNVEEDDSQDAKVLKGGLDGTYRLQFHFHWGSLDGQSEHTVDKKYAAELHVHN 124
  
```

RESULT 10
 US-08-446-363-9

Sequence 9, Application US/08446363
 Patent No. 589700

GENERAL INFORMATION:
 APPLICANT: Schlessinger, Joseph
 TITLE OF INVENTION: NOVEL RECEPTOR-TYPE PHOSPHOTYROSINE
 NUMBER OF SEQUENCES: 14
 CORRESPONDENCE ADDRESS:
 STREET: 1155 Avenue of the Americas
 CITY: New York
 STATE: New York
 COUNTRY: U.S.A.
 ZIP: 10016

COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent in Release #1.0, Version #1.25

CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/446,363
 FILING DATE:
 CLASSIFICATION: 435
 PRIORITY APPLICATION DATA:
 APPLICATION NUMBER: US 08/015,986
 FILING DATE: 10-FEB-1993
 ATTORNEY/AGENT INFORMATION:
 NAME: Misrock, S. Leslie
 REGISTRATION NUMBER: 18,872
 REFERENCE/DOCKET NUMBER: 7683-028

TELECOMMUNICATION INFORMATION:
 TELEPHONE: 212-790-9090
 TELEFAX: 212-869-8864/9741
 TELEX: 66141 PENNIE

INFORMATION FOR SEQ ID NO: 9:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 260 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: unknown
 MOLECULE TYPE: protein

US-08-446-363-9

Query Match Best Local Similarity 61.2%; Score 788.5; DB 2; Length 260;
 Matches 143; Conservative 33; Mismatches 55; Indels 1; Gaps 1;

```

  Qy   6 WGYREHNGPIHWKEFFPPIADGDOQSPIEIKTKEYVKKDSSURPLSIKYDPSSAKIISNSGH 65
  
```


CORRESPONDENCE ADDRESS:
 ADDRESSEE: Merchant & Gould
 STREET: 3100 No. 596227owest Center, 90 S. 7th Street
 CITY: Minneapolis
 STATE: MN
 COUNTRY: U.S.A.
 ZIP: 55402
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Diskette
 COMPUTER: IBM Compatible
 OPERATING SYSTEM: DOS
 SOFTWARE: FASTSEQ Version 1.5
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08
 FILING DATE: 06-FEB-1996
 CLASSIFICATION: 435
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER:
 FILING DATE:
 ATTORNEY AGENT INFORMATION:
 NAME: Carter, Charles G
 REGISTRATION NUMBER: 35,093
 TELECOMMUNICATION INFORMATION:
 REFERENCE/DOCKET NUMBER: 8648.59US01
 TELEPHONE: 612/332-5300
 TELEXFAX: 612/332-9081
 INFORMATION FOR SEQ ID NO: 12:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 288 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 HYPOTHETICAL: NO
 ANTI-SENSE: NO
 FRAGMENT TYPE: C-terminal
 ORIGINAL SOURCE:
 US-08-595-868C-12

Query Match 61.2%; Score 788.5; DB 2; Length 288;
 Best Local Similarity 61.6%; Pred. No. 9.4e-75;
 Matches 143; Conservative 33; Mismatches 55; Indels 1; Gaps 1;

Qy 6 WGYREHNGPITHKEFFPIADGDDQOSPLIEKTEIKVYKDSSLRPLSLTKYDPSAKIISNSGH 65
 Db 5 WGYKGKINGPETHWKDPIAKGRQSPVDPIDHTAKCDPSLRLNNGH 64
 Qy 66 SENVDDDTENKSVLRRGGPLTGSYRLRQVHLWGSADDHGSEHHIVDGVSIAELHVHN 125
 Db 65 AFNVEFDSDQAVLKGGPLDGTYRILQFHWGSDGQSEHTVDDKKYAAELHVHN 124
 Qy 126 SDKYPSEVAAREHPDGLAVLGFLQPEPSQLPLSITDLSIKEKGKOTRFTNFDLSS 185
 Db 125 T-KYDFGKAQQPDPGLAVLGFLKYGSAXGLQKVVDVLDSIKYGKSADEFNFPRGL 183
 Qy 186 LPPSMWDYWTYPPGSLTVPLLESVTWVLUKOPINISSQLAKEFSLCTARGE 237
 Db 184 LPESLDWYTYPGSLTTPPLBCTVWVLUKEPISVSSBQVLRKFRLNFGEGE 235

RESULT 14
 US-09-139-819A-12

Sequence 12, Application US/09139819A

Patent No. 621635

GENERAL INFORMATION:

APPLICANT: WAGNER, Fred W.

APPLICANT: STOUT, Jay S.

APPLICANT: HENRIKSEN, Dennis B.

APPLICANT: PARTRIDGE, Bruce E.

HOLMOUST, Bart

FANK, Julie A.

TITLE OF INVENTION: RECOMBINANT PREPARATION OF CALCITONIN

FRAGMENTS AND USE THEREOF IN THE PREPARATION OF CALCITONIN

AND RELATED ANALOGS

NUMBER OF SEQUENCES: 51

TITLE OF INVENTION: FRAGMENTS AND USE THEREOF IN THE PREPARATION OF CALCITONIN
 TITLE OF INVENTION: AND RELATED ANALOGS
 NUMBER OF SEQUENCES:
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: FOLEY & LARDNER
 STREET: 3000 K Street, N.W.
 CITY: Washington
 STATE: D.C.
 COUNTRY: U.S.A.
 ZIP: 20007-5109
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent In Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09/139,819A
 FILING DATE: 25-AUG-1998
 PRIORITY APPLICATION DATA:
 APPLICATION NUMBER: US 08/595,868
 FILING DATE: 06-FEB-1996
 ATTORNEY/AGENT INFORMATION:
 NAME: BENT, Stephen A.
 REGISTRATION NUMBER: 29,768
 REFERENCE/DOCKET NUMBER: 089187/0144
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (202) 672-5300
 TELEFAX: (202) 672-5399
 INFORMATION FOR SEQ ID NO: 12:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 288 amino acids
 SEQUENCE CHARACTERISTICS:
 LENGTH: 288 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-09-139-819A-12

Query Match 61.2%; Score 788.5; DB 3; Length 288;
 Best Local Similarity 61.6%; Pred. No. 9.4e-15;
 Matches 143; Conservative 33; Mismatches 55; Indels 1; Gaps 1;

Qy 6 WGYREHNGPITHKEFFPIADGDDQOSPLIEKTEIKVYKDSSLRPLSLTKYDPSAKIISNSGH 65
 Db 5 WGYKGKINGPETHWKDPIAKGRQSPVDPIDHTAKCDPSLRLNNGH 64
 Qy 126 SDKYPSEVAAREHPDGLAVLGFLQPEPSQLPLSITDLSIKEKGKOTRFTNFDLSS 185
 Db 125 T-KYDFGKAQQPDPGLAVLGFLKYGSAXGLQKVVDVLDSIKYGKSADEFNFPRGL 183
 Qy 186 LPPSMWDYWTYPPGSLTTPPLBCTVWVLUKEPISVSSBQVLRKFRLNFGEGE 237
 Db 184 LPESLDWYTYPGSLTTPPLCETWVLUKEPISVSSBQVLRKFRLNFGEGE 235

RESULT 15
 US-09-750-913-12
 Sequence 12, Application US/09750913
 Patet No. 6410707
 GENERAL INFORMATION:
 APPLICANT: WAGNER, Fred W.
 APPLICANT: STOUT, Jay S.
 APPLICANT: HENRIKSEN, Dennis B.
 APPLICANT: PARTRIDGE, Bruce E.
 APPLICANT: HOLMOUST, Bart
 APPLICANT: FANK, Julie A.
 APPLICANT: HOLMOUST, Bart
 APPLICANT: FANK, Julie A.
 TITLE OF INVENTION: RECOMBINANT PREPARATION OF CALCITONIN
 FRAGMENTS AND USE THEREOF IN THE PREPARATION OF CALCITONIN
 AND RELATED ANALOGS

CORRESPONDENCE ADDRESS:
 ADDRESSEE: FOLEY & LARDNER
 STREET: 3000 K Street, N.W.
 CITY: Washington
 STATE: D.C.
 ZIP: 20007-5109

COMPUTER READABLE FORM:

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent In Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/750,913

FILING DATE: 12-Jan-2001

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/187,0144

FILING DATE: <Unknown>

ATTORNEY/AGENT INFORMATION:

NAME: BENT, Stephen A.

REGISTRATION NUMBER: 29,768

REFERENCE/DOCKET NUMBER: 08/187,0144

TELECOMMUNICATION INFORMATION:

TELEPHONE: (202) 672-5000

TELEFAX: (202) 672-5399

INFORMATION FOR SEQ ID NO: 12:

SEQUENCE CHARACTERISTICS:

LENGTH: 288 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

SEQUENCE DESCRIPTION: SEQ ID NO: 12:

US-09-750-913-12

Query Match 61.2%; Score 788.5; DB 4; Length 288;
 Best Local Similarity 61.6%; Pred. No. 9.4e-75; Mismatches 33; Indels 1; Gaps 1;

Matches 143; Conservative 33; Mismatches 55; Indels 1; Gaps 1;

Qy 6 WGYREHNGPDIHWKEFPIDGDOOSPIEIKTKREVKYDSLRLPISTIKDPSAKIISNSGH 65

Db 5 WGIGKHNQPEWHKDFPATAKGERSPVNIDHTAKYDDSLKLPSVSYQOATSLRILNGH 64

Qy 66 SENYDFDDTENKSYLRCGPBLTGSTRYLRLTHAAGSDDHGSEEVTDGTSYAABLHVWN 125

Db 65 AFNVEFDSDQDKAVLKGGGLDGTFLQIHFHNGSLDGQSSEITVDKKYAAELHLHVWN 124

Qy 126 SDKYPSFYBAAHEPDGLAVLGVLGVFLQIGEPNSOLOKITDTDSIKEKGKOTRFITNFDLJSL 185

Db 125 T-KYGFGRKAVQQPDGLAVLGFLKVGSAKPLQKVVDLDSIKTGSAADFNFDPGL 183

Qy 186 LPPSWDWTYPPGSLTVPPLIESUTWIVLKOPINISSQQLAKERSLLCTAEGE 237

Db 184 LPESLDWTYPPGSLTTPPLECWTWIVLKEPIVSSEQVTLKFRKLNFNGEGE 235

Search completed: September 9, 2004, 15:15:23

Job time : 38 secs

| Result No. | Score | Query Match Length DB ID | Description |
|------------|-------|--------------------------|---|
| 1 | 1289 | 100.0 | Sequence 1, Appli |
| 2 | 1289 | 100.0 | Sequence 1, Appli |
| 3 | 1289 | 100.0 | Sequence 1, Appli |
| 4 | 1289 | 100.0 | Sequence 1, Appli |
| 5 | 807 | 62.6 | Sequence 1, Appli |
| 6 | 807 | 62.6 | Sequence 1, Appli |
| 7 | 804 | 62.4 | Sequence 1, Appli |
| 8 | 804 | 62.4 | Sequence 1, Appli |
| 9 | 804 | 62.4 | Sequence 1, Appli |
| 10 | 804 | 62.4 | Sequence 1, Appli |
| 11 | 804 | 62.4 | Sequence 1, Appli |
| 12 | 802 | 62.2 | Sequence 1, Appli |
| 13 | 802 | 62.2 | Sequence 1, Appli |
| 14 | 788.5 | 61.2 | Sequence 1, Appli |
| 15 | 788.5 | 61.2 | Sequence 1, Appli |
| 16 | 788.5 | 61.2 | Sequence 1, Appli |
| 17 | 788.5 | 61.2 | Sequence 1, Appli |
| 18 | 748.5 | 58.1 | Sequence 1, Appli |
| 19 | 748.5 | 58.1 | Sequence 1, Appli |
| 20 | 748 | 58.0 | Sequence 1, Appli |
| 21 | 747.5 | 58.0 | Sequence 1, Appli |
| 22 | 747.5 | 58.0 | Sequence 1, Appli |
| 23 | 695 | 53.9 | Sequence 1, Appli |
| 24 | 695 | 53.9 | Sequence 1, Appli |
| 25 | 668 | 51.8 | Sequence 1, Appli |
| 26 | 506 | 39.3 | Sequence 1, Appli |
| 27 | 453 | 35.1 | Sequence 1, Appli |
| 28 | 453 | 35.1 | Sequence 1, Appli |
| 29 | 451 | 35.0 | Sequence 1, Appli |
| 30 | 451 | 35.0 | Sequence 1, Appli |
| 31 | 451 | 35.0 | Sequence 1, Appli |
| 32 | 451 | 35.0 | Sequence 1, Appli |
| 33 | 411.5 | 31.9 | Sequence 1, Appli |
| 34 | 411.5 | 31.9 | Sequence 1, Appli |
| 35 | 411 | 31.9 | Sequence 1, Appli |
| 36 | 376.5 | 29.2 | Sequence 1, Appli |
| 37 | 376.5 | 29.2 | Sequence 1, Appli |
| 38 | 376.5 | 29.2 | Sequence 1, Appli |
| 39 | 376.5 | 29.2 | Sequence 1, Appli |
| 40 | 376.5 | 29.2 | Sequence 1, Appli |
| 41 | 376.5 | 29.2 | Sequence 1, Appli |
| 42 | 376.5 | 29.2 | Sequence 1, Appli |
| 43 | 376.5 | 29.2 | Sequence 1, Appli |
| 44 | 376.5 | 29.2 | Sequence 1, Appli |
| 45 | 376.5 | 29.2 | Sequence 1, Appli |
| | | | ALIGNMENTS |
| | | | RESULT 1 |
| | | | ; Sequence 1, Application US/10069434 |
| | | | ; GENERAL INFORMATION: |
| | | | ; APPLICANT: INCYTE GENOMICS, INC. |
| | | | ; ATTORNEY: THORTON, Michael |
| | | | ; ATTORANT: RAMKMAR, Jayalakmi |
| | | | ; ATTORANT: TRIBOUTLEY, Catherine M. |
| | | | ; ATTORANT: YUE, Henry |
| | | | ; ATTORANT: NGUYEN, Dannie B. |
| | | | ; ATTORANT: YAO, Monique G. |
| | | | ; ATTORANT: PATERSON, Chandra |
| | | | ; ATTORANT: GANDHI, Ameena R. |
| | | | ; ATTORANT: BURFORD, Neil |
| | | | ; ATTORANT: THANARVELU, Karitha |
| | | | ; ATTORANT: BAUGHN, Mariah R. |
| | | | ; TITLE OF INVENTION: HUMAN LYASES |
| | | | ; FILE REFERENCE: PT-0137.PCT |
| | | | ; CURRENT APPLICATION NUMBER: US/10/0659,434 |
| | | | ; PRIORITY APPLICATION NUMBER: 60/213,393; 60/215,544; 60/222,818 |
| | | | ; CURRENT FILING DATE: 2002-02-20 |
| | | | ; PRIORITY FILING DATE: 2000-06-23; 2000-06-30; 2000-08-04 |
| | | | ; NUMBER OF SEQ ID NOS: 6 |
| | | | ; SEQ ID NO: 1 |
| | | | ; LENGTH: 242 |
| | | | ; TYPE: PRT |
| | | | ; ORGANISM: Homo sapiens |
| | | | ; FEATURE: |
| | | | ; NAME/KEY: misc feature |
| | | | ; OTHER INFORMATION: Incyte ID No. US20030121061A1 6338333CD1 |
| | | | US-10-069-434-1 |
| | | | Query Match Score 100.0%; Score 1289; DB 14; Length 242; |
| | | | Best Local Similarity 100.0%; Pred. No. 3 9e-125; |
| | | | Matches 242; Conservative 0; Mismatches 0; Indels 0; Gaps 0; |

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match Length DB ID | Description |
|------------|-------|--------------------------|-------------------|
| 1 | 1289 | 100.0 | Sequence 1, Appli |
| 2 | 1289 | 100.0 | Sequence 1, Appli |
| 3 | 1289 | 100.0 | Sequence 1, Appli |
| 4 | 1289 | 100.0 | Sequence 1, Appli |
| 5 | 807 | 62.6 | Sequence 1, Appli |
| 6 | 807 | 62.6 | Sequence 1, Appli |
| 7 | 804 | 62.4 | Sequence 1, Appli |
| 8 | 804 | 62.4 | Sequence 1, Appli |
| 9 | 804 | 62.4 | Sequence 1, Appli |
| 10 | 804 | 62.4 | Sequence 1, Appli |
| 11 | 804 | 62.4 | Sequence 1, Appli |
| 12 | 802 | 62.2 | Sequence 1, Appli |
| 13 | 802 | 62.2 | Sequence 1, Appli |
| 14 | 788.5 | 61.2 | Sequence 1, Appli |
| 15 | 788.5 | 61.2 | Sequence 1, Appli |

Qy 1 MSRLSNGYREHNGPITHKEFFPIADGDOQSPIEIKTKEVYKDYSSLRPLSIKYDPSSAKII 60
 Db 1 MSRLSNGYREHNGPITHKEFFPIADGDOQSPIEIKTKEVYKDYSSLRPLSIKYDPSSAKII 60

Qy 61 SNSGHSFNVDFDDTENKSVLRGGLTGSPYLGSYRLRQVHLHWGSAADDHGSEEHIVDGVSAAELH 120
 Db 61 SNSGHSFNVDFDDTENKSVLRGGLTGSPYLGSYRLRQVHLHWGSAADDHGSEEHIVDGVSAAELH 120

Qy 121 VWHWNSDKYPSVEAAHEPDGLAVLGVLQIGEPNSLQOKITDTDSIKEKGKQTTRTNF 180
 Db 121 VWHWNSDKYPSVEAAHEPDGLAVLGVLQIGEPNSLQOKITDTDSIKEKGKQTTRTNF 180

Qy 181 DLISLLPPSWDWTYPCGSLLTVPPLESVTWIVLKQPINISSQQLAKERSLCTAEGEAAA 240
 Db 181 DLISLLPPSWDWTYPCGSLLTVPPLESVTWIVLKQPINISSQQLAKERSLCTAEGEAAA 240

Qy 241 FL 242
 Db 241 FL 242

Qy 241 FL 242
 Db 241 FL 242

RESULT 3
 US-10-108-260A-3082
 ; Sequence 3082, Application US/10108260A
 ; Publication No. US200400566A1
 ; GENERAL INFORMATION:
 ; APPLICANT: HELIX RESEARCH INSTITUTE
 ; TITLE OF INVENTION: NO. US200400566A1 full length cDNA
 ; FILE REFERENCE: H1-A0106
 ; CURRENT APPLICATION NUMBER: US/10/108-260A
 ; CURRENT FILING DATE: 2002-03-27
 ; NUMBER OF SEQ ID NOS: 5458
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO: 3082
 ; LENGTH: 262
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 ; US-10-108-260A-3082

Query Match 100.0%; Score 1289; DB 15; Length 262;
 Best Local Similarity 100.0%; Pred. No. 4-e-125;
 Matches 242; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MSRLSNGYREHNGPITHKEFFPIADGDOQSPIEIKTKEVYKDYSSLRPLSIKYDPSSAKII 60
 Db 1 MSRLSNGYREHNGPITHKEFFPIADGDOQSPIEIKTKEVYKDYSSLRPLSIKYDPSSAKII 60

Qy 61 SNSGHSFNVDFDDTENKSVLRGGLTGSPYLGSYRLRQVHLHWGSAADDHGSEEHIVDGVSAAELH 120
 Db 61 SNSGHSFNVDFDDTENKSVLRGGLTGSPYLGSYRLRQVHLHWGSAADDHGSEEHIVDGVSAAELH 120

Qy 121 VWHWNSDKYPSVEAAHEPDGLAVLGVLQIGEPNSLQOKITDTDSIKEKGKQTTRTNF 180
 Db 121 VWHWNSDKYPSVEAAHEPDGLAVLGVLQIGEPNSLQOKITDTDSIKEKGKQTTRTNF 180

Qy 181 DLISLLPPSWDWTYPCGSLLTVPPLESVTWIVLKQPINISSQQLAKERSLCTAEGEAAA 240
 Db 181 DLISLLPPSWDWTYPCGSLLTVPPLESVTWIVLKQPINISSQQLAKERSLCTAEGEAAA 240

Qy 241 FL 242
 Db 241 FL 242

RESULT 4
 US-10-240-120-215
 ; Sequence 215, Application US/10220120
 ; Publication No. US20040048253A1
 ; GENERAL INFORMATION:
 ; APPLICANT: INCYTE GENOMICS, INC.
 ; PANZER, Scott R.
 ; SPIRO, Peter A.
 ; BANVILLE, Steven C.

Query Match 100.0%; Score 1289; DB 12; Length 262;
 Best Local Similarity 100.0%; Pred. No. 4-e-125;
 Matches 242; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

US-10-433-802-1

; Sequence 1, Application US/10433802
 ; Publication No. US20040063115A1
 ; GENERAL INFORMATION:
 ; APPLICANT: TANG, Y.; TOM, GRIFFIN, Jennifer A.;
 ; APPLICANT: YUE, Henry; LEE, Ernestine A.;
 ; APPLICANT: BAUGHN, Mariah R.; DUGAN, Brendan M.;
 ; APPLICANT: CHAWLA, Narinder K.; LEE, Sally;
 ; APPLICANT: RAMKUMAR, Jayalaxmi; WARREN, Bridget A.;
 ; APPLICANT: GANDHI, Ameena R.; LU, Dyring Alina M.;
 ; APPLICANT: LIU, Yan; YAO, Monique G.;
 ; APPLICANT: DING, Li; TRIBOURY, Catherine M.;
 ; APPLICANT: SANJAYALA, Madhusudan M.;
 ; APPLICANT: ARVIZI, Chandri S.; JACKSON, Jennifer L.
 ; TITLE OF INVENTION: ENZYME
 ; FILE REFERENCE: PI-0316 PCT
 ; CURRENT APPLICATION NUMBER: US/10/433,902
 ; CURRENT FILING DATE: 2003-06-04
 ; PRIOR APPLICATION NUMBER: PCT/US01/47432
 ; PRIOR FILING DATE: 2001-12-04
 ; PRIOR APPLICATION NUMBER: US 60/251,824
 ; PRIOR FILING DATE: 2000-12-07
 ; PRIOR APPLICATION NUMBER: US 60/254,312
 ; PRIOR FILING DATE: 2000-12-08;
 ; PRIOR APPLICATION NUMBER: US 60/255,773
 ; PRIOR FILING DATE: 2000-12-14
 ; PRIOR APPLICATION NUMBER: US 60/256,188
 ; PRIOR FILING DATE: 2000-12-15
 ; PRIOR APPLICATION NUMBER: US 60/255,940
 ; PRIOR FILING DATE: 2000-12-15
 ; PRIOR APPLICATION NUMBER: US 60/257,488
 ; PRIOR FILING DATE: 2000-12-21
 ; PRIOR APPLICATION NUMBER: US 60/262,839
 ; PRIOR FILING DATE: 2001-01-19
 ; PRIOR APPLICATION NUMBER: US 60/264,402
 ; PRIOR FILING DATE: 2001-01-26
 ; NUMBER OF SEQ ID NOS: 36
 ; SOFTWARE: PERL Program
 ; SEQ ID NO: 1
 ; LENGTH: 262
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; NAME/KEY: misc feature
 ; OTHER INFORMATION: Incyte ID No. US20040063115A1 8159895CD1
 ; US-10-433-802-1

Publication No. US20040018196A1
 GENERAL INFORMATION:
 APPLICANT: Mezes et al.
 TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same
 FILE REFERENCE: 21402-240
 CURRENT APPLICATION NUMBER: US/10/044,564
 CURRENT FILING DATE: 2002-05-09
 PRIORITY APPLICATION NUMBER: 60/261,014
 PRIOR FILING DATE: 2001-01-11
 PRIOR APPLICATION NUMBER: 60/261,018
 PRIOR FILING DATE: 2001-01-11
 PRIORITY APPLICATION NUMBER: 60/318,410
 PRIOR FILING DATE: 2001-09-10
 PRIORITY APPLICATION NUMBER: 60/261,013
 PRIOR FILING DATE: 2001-01-11
 PRIOR APPLICATION NUMBER: 60/261,026
 PRIOR FILING DATE: 2001-01-11
 PRIOR APPLICATION NUMBER: 60/261,029
 PRIOR FILING DATE: 2001-01-11
 PRIORITY APPLICATION NUMBER: 60/313,170
 PRIOR FILING DATE: 2001-08-17
 NUMBER OF SEQ ID NOS: 306
 SOFTWARE: PatentIn Ver. 2.1
 SEQ ID NO: 304
 LENGTH: 255
 TYPE: PRT
 ORGANISM: Artificial Sequence
 FEATURE:
 OTHER INFORMATION: Description of Artificial Sequence:consensus
 US-10-044-564-304

Query Match 62.4% Score 804; DB 9; Length 261;
 Best Local Similarity 59.4%; Pred. No. 9.7e-75;
 Matches 142; Conservative 44; Mismatches 53; Indels 0; Gaps 0;

Qy 1 MSRLSWGYREHNGPIHWKEFFPIADGDOQSPRIKTRKVYDSSLRPLSIKYDPSAKTI 60
 Db 1 MASPDWGYDDRGPEQNSKLYPIANGNQSPIDKTSETKHTSLKPSVSYNPATAKEI 60

Qy 61 SNSGHSENVDDETDENKSVLRGGLPTGSYRLRQVHLWGSADDHGSEHIVDGYSAAELH 120
 Db 61 INVGHSHFVNFDNDNRSVLKGGPFSYRLFQFHFMGSTNEHGSBHTDVQKYSAEH 120

Qy 121 VVHWNSDKYSPFVEAAHEPDGLAVLGVLQIGEPNSLQKTTDLSIKEKGKQTRETFNF 180
 Db 121 VAHWNSAKYSSLAEAASKADGLAVLGVLQKVKVLDQAIKTKGKRAPTFNF 180

Qy 181 DILSLPPSDWYTYPGSLTVPPLESVTWIVKQPINISSQLAKEPSLLCTAGEAA 239
 Db 181 DPSTILLPSLDFWTYPPSLLTHPPLESVTWIVKQPINISSQLAKEPSLLCTAGEAA 239

RESULT 8
 US-09-981-353-80
 ; Sequence 80, Application US/09981353
 ; Patent No. US2010160382A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Lasek, Amy W.
 ; APPLICANT: Jones, David A.
 ; TITLE OF INVENTION: GENES EXPRESSED IN COLON CANCER
 ; FILE REFERENCE: PA-0038 US
 ; CURRENT APPLICATION NUMBER: US/09/981,353
 ; CURRENT FILING DATE: 2001-10-11
 ; NUMBER OF SEQ ID NOS: 194
 ; SEQ ID NO: 80
 ; LENGTH: 261
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; NAME/KEY: misc_feature
 ; OTHER INFORMATION: Incyte ID No. US20020160382A1 210163CD1
 US-09-981-353-80

Query Match 62.4% Score 804; DB 9; Length 261;
 Best Local Similarity 59.4%; Pred. No. 9.7e-75;
 Matches 142; Conservative 44; Mismatches 53; Indels 0; Gaps 0;

Qy 1 MSRLSWGYREHNGPIHWKEFFPIADGDOQSPRIKTRKVYDSSLRPLSIKYDPSAKTI 60
 Db 1 MASPDWGYDDRGPEQNSKLYPIANGNQSPIDKTSETKHTSLKPSVSYNPATAKEI 60

Qy 61 SNSGHSENVDDETDENKSVLRGGLPTGSYRLRQVHLWGSADDHGSEHIVDGYSAAELH 120
 Db 61 INVGHSHFVNFDNDNRSVLKGGPFSYRLFQFHFMGSTNEHGSBHTDVQKYSAEH 120

Qy 121 VVHWNSDKYSPFVEAAHEPDGLAVLGVLQIGEPNSLQKTTDLSIKEKGKQTRETFNF 180
 Db 121 VAHWNSAKYSSLAEAASKADGLAVLGVLQKVKVLDQAIKTKGKRAPTFNF 180

Qy 181 DILSLPPSDWYTYPGSLTVPPLESVTWIVKQPINISSQLAKEPSLLCTAGEAA 239
 Db 181 DPSTILLPSLDFWTYPPSLLTHPPLESVTWIVKQPINISSQLAKEPSLLCTAGEAA 239

RESULT 9
 US-10-235-994-24
 ; Sequence 24, Application US/10235994
 ; Publication No. US20030101002A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Macina, Roberto A.
 ; APPLICANT: Paderit, Alejandra
 ; APPLICANT: Sun, Yongming
 ; TITLE OF INVENTION: Method of Diagnosing, Monitoring, Staging, Imaging and
 ; TITLE OF INVENTION: Treating Gastrointestinal cancer
 ; FILE REFERENCE: DE-0142
 ; CURRENT FILING DATE: 2001-03-09
 ; PRIORITY APPLICATION NUMBER: US/09/802,674
 ; PRIOR FILING DATE: 2000-03-09
 ; NUMBER OF SEQ ID NOS: 13
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO: 4
 ; LENGTH: 261
 ; TYPE: PRT

i APPLICANT: Walker, Michael
i TITLE OF INVENTION: METHODS FOR ANALYZING GENE EXPRESSION PATTERNS
i FILE REFERENCE: ICYTF012
i CURRENT APPLICATION NUMBER: US/10/235,994
i CURRENT FILING DATE: 2002-09-04
i PRIOR APPLICATION NUMBER: US/10/003,608
i PRIOR FILING DATE: 2001-11-01
i PRIOR FILING DATE: 2000-11-01
i NUMBER OF SEQ ID NOS: 30
i SOFTWARE: FastSEQ for Windows Version 4.0
i SEQ ID NO: 24
i LENGTH: 261
i TYPE: PRT
i ORGANISM: Human
US-10-235-994-24

Query Match 62.4%; Score 804; DB 14; Length 261;
Best Local Similarity 59.4%; Pred. No. 9.7e-75; Indels 0; Gaps 0;
Matches 142; Conservative 44; Mismatches 53; Indels 0; Gaps 0;
Qy 1 MSRLSWGYRBRHNGPIHWKEFFPIAHDQDQSPIEIKTKKEYDSSRLPLSKYDPSSAKII 60
Db 1 MASPDWGYDDKNGPQWSKLYPIANGNNQSPVLDKXSETHTDKLPSKISYNPAKAEI 60
Qy 61 SNSGHSFNVDDTENKSVLRRGGPLTSYRLRQVHLWGSADDHGSEHTWDGVSYAAELH 120
Db 61 INVGHSHPNEDNDRSVPKGPFSDSYLFLQFIFHMGSTNEHGEHTWDGVSYAAELH 120
Qy 121 VVHNSDKYPSEVEAAHPEGCLAVLGVLQIGEPNSQLOKITDLSIKEKGKOTRFNF 180
Db 121 VAHNSAKYSSLAEEASKADGLAVGVLMRVGAEANPKLQKVLDQAIKTKGRAPFTNF 180
Qy 181 DLLSLPPSDYWTYPPGSLTVPPPLESVTIVLKOPINISSQQLAKFRSLLCTAGEAA 239
Db 181 DPSTLPPSSLDFWTYPPGSLTHPPLEXSVTIVICKESISVSSEQLAQFRSLLSNVGDNA 239
RESULT 10
US-10-408-765A-389
i Sequence 389, Application US/10/0408765A
i Publication No. US20040101874A1
i GENERAL INFORMATION:
i APPLICANT: Ghosh, Sounmitra S.
i APPLICANT: Fahy, John D.
i APPLICANT: Zhang, Bing
i APPLICANT: Gibson, Bradford W.
i APPLICANT: Taylor, Steven W.
i APPLICANT: Gleim, Gary M.
i APPLICANT: Warnock, Dale E.
i TITLE OF INVENTION: TARGETS FOR THERAPEUTIC INTERVENTION
i TITLE OF INVENTION: IDENTIFIED IN THE MITOCHONDRIAL PROTEOME
i FILE REFERENCE: 66008.465
i CURRENT APPLICATION NUMBER: US/10/408,765A
i CURRENT FILING DATE: 2003-04-04
i NUMBER OF SEQ ID NOS: 3077
i SOFTWARE: FastSEQ for Windows Version 4.0
i SEQ ID NO: 389
i LENGTH: 261
i TYPE: PRT
i ORGANISM: Homo sapiens
US-10-408-765A-389

Query Match 62.4%; Score 804; DB 16; Length 261;
Best Local Similarity 59.4%; Pred. No. 9.7e-75; Indels 0; Gaps 0;
Matches 142; Conservative 44; Mismatches 53; Indels 0; Gaps 0;
Qy 1 MSRLSWGYRBRHNGPIHWKEFFPIAHDQDQSPIEIKTKKEYDSSRLPLSKYDPSSAKII 60
Db 1 MASPDWGYDDKNGPQWSKLYPIANGNNQSPVLDKXSETHTDKLPSKISYNPAKAEI 60
Qy 61 SNSGHSFNVDDTENKSVLRRGGPLTSYRLRQVHLWGSADDHGSEHTWDGVSYAAELH 120
Db 63 INVGHSHPNEDNDRSVPKGPFSDSYLFLQFIFHMGSTNEHGEHTWDGVSYAAELH 120
Qy 121 VVHNSDKYPSEVEAAHPEGCLAVLGVLQIGEPNSQLOKITDLSIKEKGKOTRFNF 180
Db 123 VAHNSAKYSSLAEEASKADGLAVGVLMRVGAEANPKLQKVLDQAIKTKGRAPFTNF 182
RESULT 12
US-09-983-0004-26
i Sequence 26, Application US/09983000A
i Publication No. US20030118585A1
i GENERAL INFORMATION:
i APPLICANT: AGY Therapeutics
i APPLICANT: Melcher, Thorsten
i APPLICANT: Mueller, Sabine
i APPLICANT: Chin, Daniel
i TITLE OF INVENTION: USE OF PROTEIN BIOMOLECULAR TARGETS IN THE TREATMENT AND VISUALIZA
i FILE REFERENCE: 263/180 -- Peagleman -- AGY
i CURRENT APPLICATION NUMBER: US/09/983,000A
i CURRENT FILING DATE: 2001-10-17
i NUMBER OF SEQ ID NOS: 35
i SOFTWARE: PatentIn version 3.1
i SEQ ID NO: 26

Query Match 62.4%; Score 804; DB 14; Length 263;
Best Local Similarity 59.4%; Pred. No. 9.8e-75; Indels 0; Gaps 0;
Matches 142; Conservative 44; Mismatches 53; Indels 0; Gaps 0;
Qy 1 MSRLSWGYRBRHNGPIHWKEFFPIAHDQDQSPIEIKTKKEYDSSRLPLSKYDPSSAKII 60
Db 3 MASPDWGYDDKNGPQWSKLYPIANGNNQSPVLDKXSETHTDKLPSKISYNPAKAEI 62
Qy 61 SNSGHSFNVDDTENKSVLRRGGPLTSYRLRQVHLWGSADDHGSEHTWDGVSYAAELH 120
Db 63 INVGHSHPNEDNDRSVPKGPFSDSYLFLQFIFHMGSTNEHGEHTWDGVSYAAELH 120
Qy 121 VVHNSDKYPSEVEAAHPEGCLAVLGVLQIGEPNSQLOKITDLSIKEKGKOTRFNF 180
Db 123 VAHNSAKYSSLAEEASKADGLAVGVLMRVGAEANPKLQKVLDQAIKTKGRAPFTNF 182
RESULT 11
US-10-106-698-4637
i Sequence 4637, Application US/10106698
i Publication No. US20031010690A1
i GENERAL INFORMATION:
i APPLICANT: Ruben et al.
i TITLE OF INVENTION: Colon and Colon Cancer Associated Polynucleotides and Polypeptides
i FILE REFERENCE: PA005P1
i CURRENT APPLICATION NUMBER: US/10/106,638
i CURRENT FILING DATE: 2002-03-27
i PRIORITY APPLICATION NUMBER: PCT/US00/26524
i PRIORITY FILING DATE: 2000-09-28
i PRIORITY APPLICATION NUMBER: US 60/157,137
i PRIORITY FILING DATE: 1999-05-29
i PRIORITY APPLICATION NUMBER: US 60/163,280
i PRIORITY FILING DATE: 1999-11-03
i NUMBER OF SEQ ID NOS: 8564
i SOFTWARE: PatentIn Ver. 3.0
i SEQ ID NO: 4637
i TYPE: PRT
i ORGANISM: Homo sapiens
US-10-106-698-4637

LENGTH: 260
 TYPE: PRF
 ORGANISM: Homo sapiens
 FEATURE: Gene
 NAME/KEY: Gene
 LOCATION: (1)..(260)
 OTHER INFORMATION: Carbonic anhydrase domain of human carbonic anhydrase I
 US-09-983-00A-26

Query Match 62.2%; Score 802; DB 10; Length 260;
 Best Local Similarity 60.3%; Pred. No. 1..6-74;
 Matches 141; Conservative 43; Mismatches 50; Indels 0; Gaps 0;

Qy 6 WGYREHNGPIHWKEFPPIADGQQSPLIEKTKEVKXDSLRPLSIKAKLISNSGH 65
 5 WGYDDKNGPEQWKLPIANGNQSPVDIKTSETKIDTSRKPSVSYNPAKETINVGH 64

Db 6 SFNVDFDDTENKSVLRGSPLGPSYRDRQHJHGSAADDHGSEHIVDGVSYAAELHVHN 125
 5 SFHVNFEDNDRSVLXKGPFDSYRLFQFHFWGSTNEHGSEHTVDGVYSAELHVHN 124

Qy 65 SFHVNFEDNDRSVLXKGPFDSYRLFQFHFWGSTNEHGSEHTVDGVYSAELHVHN 124

Db 126 SDKYPSEVEAAHEPDGUAVLGVLGFQLOGEPSNOLOKITDTDLDSIKEKGKOTRFTNFDLSL 185
 125 SAKYSSLAEEASKADGLAVIGLKVQQLKQLDALKIQTGKRAFTNFDSTL 184

Qy 126 SDKYPSEVEAAHEPDGUAVLGVLGFQLOGEPSNOLOKITDTDLDSIKEKGKOTRFTNFDLSL 185
 125 SAKYSSLAEEASKADGLAVIGLKVQQLKQLDALKIQTGKRAFTNFDSTL 184

Db 186 LPPSWDYWTYPSLITVPPLESVTWVILKOPINISSQQLAKERSLIICTAEGBAA 239
 185 LPPSWDYWTYPSLITVPPLESVTWVILKOPINISSQQLAKERSLIICTAEGBAA 239

Qy 186 LPPSWDYWTYPSLITVPPLESVTWVILKOPINISSQQLAKERSLIICTAEGBAA 239
 185 LPPSWDYWTYPSLITVPPLESVTWVILKOPINISSQQLAKERSLIICTAEGBAA 239

Db 185 LPPSWDYWTYPSLITVPPLESVTWVILKOPINISSQQLAKERSLIICTAEGBAA 239

RESULT 14
 US-10-00-954-5
 Sequence 5, Application US/10000954
 Publication No. US2002012722A1
 GENERAL INFORMATION:
 APPLICANT: Schlessinger, Joseph
 Barnea, Gilad
 Grumet, Martin H.
 Margolis, Richard U.
 TITLE OF INVENTION: A NEW CLASS OF RPTPases: THEIR
 NUMBER OF SEQUENCES: 13
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: PENNIE & EDMONDS
 STREET: 115 Avenue of the Americas
 CITY: New York
 STATE: New York
 COUNTRY: U.S.A.
 ZIP: 10036
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US10/000,954
 FILING DATE: 04-Dec-2001
 CLASSIFICATION: <Unknown>
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 09/644,293
 FILING DATE: 13-Aug-2000
 APPLICATION NUMBER: 08/081,929
 FILING DATE: <Unknown>
 ATTORNEY/AGENT INFORMATION:
 NAME: Coruzzi, Laura A.
 REGISTRATION NUMBER: 30742
 REFERENCE/DOCKET NUMBER: 7683-041-999
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 212 790-9090
 TELEX: 66141 PENNIE
 INFORMATION FOR SEQ ID NO: 4:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 259 amino acids

LENGTH: 260 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: unknown
 MOLECULE TYPE: protein
 SEQUENCE DESCRIPTION: SEQ ID NO: 4:
 US-10-000-954-4

Query Match 62.2%; Score 802; DB 13; Length 260;
 Best Local Similarity 60.3%; Pred. No. 1..6-74;
 Matches 141; Conservative 43; Mismatches 50; Indels 0; Gaps 0;

Qy 6 WGYREHNGPIHWKEFPPIADGQQSPLIEKTKEVKXDSLRPLSIKAKLISNSGH 65
 5 WGYDDKNGPEQWKLPIANGNQSPVDIKTSETKIDTSRKPSVSYNPAKETINVGH 64

Db 66 SFNVDFDDTENKSVLRGSPLGPSYRDRQHJHGSAADDHGSEHIVDGVSYAAELHVHN 125
 65 SFHVNFEDNDRSVLXKGPFDSYRLFQFHFWGSTNEHGSEHTVDGVYSAELHVHN 124

Qy 126 SDKYPSEVEAAHEPDGUAVLGVLGFQLOGEPSNOLOKITDTDLDSIKEKGKOTRFTNFDLSL 185
 125 SAKYSSLAEEASKADGLAVIGLKVQQLKQLDALKIQTGKRAFTNFDSTL 184

Db 186 LPPSWDYWTYPSLITVPPLESVTWVILKOPINISSQQLAKERSLIICTAEGBAA 239
 185 LPPSWDYWTYPSLITVPPLESVTWVILKOPINISSQQLAKERSLIICTAEGBAA 239

RESULT 14
 US-10-00-954-5
 Sequence 5, Application US/10000954
 Publication No. US2002012722A1
 GENERAL INFORMATION:
 APPLICANT: Schlessinger, Joseph
 Barnea, Gilad
 Grumet, Martin H.
 Margolis, Richard U.
 TITLE OF INVENTION: A NEW CLASS OF RPTPases: THEIR
 NUMBER OF SEQUENCES: 13
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: PENNIE & EDMONDS
 STREET: 115 Avenue of the Americas
 CITY: New York
 STATE: New York
 COUNTRY: U.S.A.
 ZIP: 10036
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US10/000,954
 FILING DATE: 04-Dec-2001
 CLASSIFICATION: <Unknown>
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 09/644,293
 FILING DATE: 13-Aug-2000
 APPLICATION NUMBER: 08/081,929
 FILING DATE: <Unknown>
 ATTORNEY/AGENT INFORMATION:
 NAME: Coruzzi, Laura A.
 REGISTRATION NUMBER: 30742
 REFERENCE/DOCKET NUMBER: 7683-041-999
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 212 790-9090
 TELEX: 66141 PENNIE
 INFORMATION FOR SEQ ID NO: 5:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 259 amino acids

TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: unknown
 MOLECULE TYPE: protein
 SEQUENCE DESCRIPTION: SEQ ID NO: 5:
 US-10-000-954-5

Query Match 61.2%; Score 788.5; DB 13; Length 259;
 Best Local Similarity 61.3%; Pred. No. 3.9e-13;
 Matches 143; Conservative 33; Mismatches 55; Indels 1; Gaps 1;

Qy 6 WGYREHNGPILWKEFPPIDGQQSPLIEIYKTEVKYKDSSRLPLSIKYDSSAKLISNSGH 65
 Db 4 WGYKRNHGPEWKHDKDPIAKGERQSPVDDHTAKYDPSLKLPLSYDQATSLRILNNGH 63

Qy 66 SENVDPDDTENKSVLRGGPILTCYSYRLQVHLHWGSADDHGSEHIVDGVSAAELHVHN 125
 Db 64 AFNVEDDSQDAKVLKGGLPDTYRLQFHFWGSLDQGSSEHTVDKKKRAELHVHN 123

Qy 126 SDKYPSPVEAAHEPDGLAFLVGLQIGEPNSLOQKTTDLSIKEKGKOTRTNFEDLLS 185
 Db 124 T-KYGDGKAYQQDPGEAVLGFLKYGSAEGLQKVVDVLSIKTKGKSADETFNFDPRGL 182

Qy 186 LPPSWDYWTYPSLITVPPLLESVTWVLKOPINISSQLAKERSLCTAGE 237
 Db 183 LPESLDYWTYPSLITPPPLECYTVLWKEPVSVSSBQVLFERKLNFNGSE 234

RESULT 15
 US-09-981-353-117
 Sequence 117, Application US/09981353
 Patent No. US20020160382A1
 GENERAL INFORMATION:
 APPLICANT: Lasek, Amy W.
 ATTORNEY OR AGENT FOR APPLICANT: Jones, David A.
 TITLE OF INVENTION: GENES EXPRESSED IN COLON CANCER
 FILE REFERENCE: PA-0038 US
 CURRENT APPLICATION NUMBER: US/09/981,353
 CURRENT FILING DATE: 2001-10-11
 NUMBER OF SEQ ID NOS: 194
 SOFTWARE: PERL Program
 SEQ ID NO: 117
 LENGTH: 260
 TYPE: PRT
 ORGANISM: Homo sapiens
 FEATURES:
 NAME/KEY: misc feature
 OTHER INFORMATION: Incyte ID NO. US20020160382A1 2742913CD1
 US-09-981-353-117

Query Match 61.2%; Score 788.5; DB 9; Length 260;
 Best Local Similarity 61.6%; Pred. No. 3.9e-13;
 Matches 143; Conservative 33; Mismatches 55; Indels 1; Gaps 1;

Qy 6 WGYREHNGPILWKEFPPIDGQQSPLIEIYKTEVKYKDSSRLPLSIKYDSSAKLISNSGH 65
 Db 5 WGYKRNHGPEWKHDKDPIAKGERQSPVDDHTAKYDPSLKLPLSYDQATSLRILNNGH 64

Qy 66 SENVDPDDTENKSVLRGGPILTCYSYRLQVHLHWGSADDHGSEHIVDGVSAAELHVHN 125
 Db 65 AFNVEDDSQDAKVLKGGLPDTYRLQFHFWGSLDQGSSEHTVDKKKRAELHVHN 124

Qy 126 SDKYPSPVEAAHEPDGLAFLVGLQIGEPNSLOQKTTDLSIKEKGKOTRTNFEDLLS 185
 Db 125 T-KYGDGKAYQQDPGEAVLGFLKYGSAEGLQKVVDVLSIKTKGKSADETFNFDPRGL 183

Qy 186 LPPSWDYWTYPSLITVPPLLESVTWVLKOPINISSQLAKERSLCTAGE 237
 Db 184 LPESLDYWTYPSLITPPPLECYTVLWKEPVSVSSBQVLFERKLNFNGSE 235

Blank
Sheet